

# **USE OF MACHINE LEARNING IN CLIMATE ECONOMETRICS**

By Sebastian Mathias Jensen

A PhD thesis submitted to  
School of Business and Social Sciences, Aarhus University,  
in partial fulfilment of the requirements of  
the PhD degree in  
Economics and Business Economics

August 2021



This version: November 8, 2021 © Sebastian Mathias Jensen

## PREFACE

This dissertation was written during the period from September 2018 to August 2021 and represents the tangible part of my graduate studies at the Department of Economics and Business Economics at Aarhus University. I am very grateful to the department and the Center for Research in Econometric Analysis of Time Series (CREATES) for providing a truly inspiring research environment. I also acknowledge financial support from the DFF grant Econometric Modeling of Climate Change.

First and foremost, I would like to thank my main supervisor Professor Eric Hillebrand. I am very grateful that you have introduced me to the field of climate econometrics. Your passion for the field is contagious. I thank you for your guidance and for always being available. I would also like to thank my co-supervisor Associate Professor Mikkel Bennedsen. You have always been eager to help with everything from RejsUD to MATLAB. You have shown much more interest in me and my work than one could expect. I thank you for showing me the ropes. The three of us have had many interesting and stimulating conversations. Chapters 1 and 2 of this dissertation represent the tangible outcome. I look forward to continue our work together in the future. I also thank Solveig Nygaard Sørensen for the proof reading and all practical matters.

During the spring of 2021, I had the pleasure of virtually visiting Professor Siem Jan Koopman at Vrije Universiteit Amsterdam. I am grateful for the insightful comments on my work with Eric and Mikkel and for helping arrange for me to give a virtual seminar presentation. Circumstances made it so that I could not travel to Amsterdam. I hope to get the chance to come and visit in the future.

I am of course also thankful to my fellow PhD students and colleagues at the department for creating an outstanding academic and social environment. I am especially grateful to past and present office mates for inspiring conversations and much needed coffee breaks: Mathias, Mads, Benjamin, Anine, Frederik, and Kenneth. Special thank you to my longest standing office mate Mathias for countless talks about machine learning, golf, and everything in between. I hope they never stop. Another special thank you to fellow PhD student, past roommate, and all around good guy Cliff. I look back and smile at the nights we spend watching "Floribama Beach" after long hours at the office. I am also thankful to my old study mates Mengsted, BV, Nielsen, and of course Team Extreme, mostly so for the times we spend together

outside of the study halls and within Fuglsangssalen and Samfundsfraglig Kantine. In addition to friends and colleagues from academia, I would like to thank my friends from outside of academia for reminding me that simulation studies and consistency results are not everything.

Finally, I have reserved a very special thank you for my family: Mor, Far, Frederik, Morten, Ida, and Camilla. Thank you for your unconditional love and support. Thank you for your understanding and your encouragement. Thank you.

*Sebastian Mathias Jensen  
Aarhus, August 2021*

## **UPDATED PREFACE**

The pre-defense was held on October 27, 2021 in Aarhus. I am grateful to the members of the assessment committee Professor Marcelo Medeiros (Pontifical Catholic University of Rio de Janeiro), Professor Zac Miller (University of Missouri), and the chair of the assessment committee Associate Professor Allan Würtz (Aarhus University) for their careful reading of the dissertation and their many insightful comments and suggestions. Some of the suggestions have been incorporated into the present version of the dissertation while others remain for future research.

*Sebastian Mathias Jensen  
Aarhus, November 2021*



# CONTENTS

<b>Summary</b>	vii
<b>Danish summary</b>	xiii
<b>1 A Neural Network Approach to the Environmental Kuznets Curve</b>	1
1.1 Introduction . . . . .	2
1.2 The Environmental Kuznets Curve and its Econometric Issues . . . . .	4
1.3 Data . . . . .	7
1.4 Methodology . . . . .	9
1.5 Monte Carlo Experiment . . . . .	18
1.6 Empirical Analysis . . . . .	31
1.7 Conclusion . . . . .	44
1.8 References . . . . .	47
Appendix . . . . .	54
<b>2 Apocalypse Now? Projecting CO<sub>2</sub> Emissions with Neural Networks</b>	69
2.1 Introduction . . . . .	70
2.2 Scenario-based emissions projections and econometric issues . . . . .	72
2.3 Data . . . . .	75
2.4 Methodology . . . . .	81
2.5 Out-of-sample experiment . . . . .	91
2.6 SSP projections . . . . .	95
2.7 Conclusion . . . . .	103
2.8 References . . . . .	105
Appendix . . . . .	112
<b>3 Nowcasting U.S. CO<sub>2</sub> Emissions using Machine Learning</b>	123
3.1 Introduction . . . . .	124
3.2 Related literature . . . . .	125
3.3 The nowcasting problem . . . . .	128
3.4 Methodology . . . . .	131
3.5 Nowcasting experiment . . . . .	143
3.6 Conclusion . . . . .	157

3.7 References . . . . .	159
Appendix . . . . .	166

## SUMMARY

This dissertation consists of three self-contained chapters on the use of machine learning in climate econometrics and is particularly concerned with how tools and ideas from the fields of econometrics and machine learning can be combined to shed new light on the relationship between macroeconomic activity and carbon dioxide (CO<sub>2</sub>) emissions. According to the Intergovernmental Panel on Climate Change (IPCC) of the United Nations, CO<sub>2</sub> emissions constitute the key driver of climate change and are driven largely by economic and population growth [IPCC] 2014], highlighting the importance of a sound understanding of the relationship between macroeconomic activity and CO<sub>2</sub> emissions.

**Chapter 1, A Neural Network Approach to the Environmental Kuznets Curve** (joint work with Professor Eric Hillebrand and Associate Professor Mikkel Bennedsen), considers gross domestic product (GDP) as an aggregated measure of macroeconomic activity and uses national-level panel data to investigate the shape of the relationship between per capita GDP and per capita CO<sub>2</sub> emissions. It is important to the public and policy makers alike to have a solid grasp of the shape of this relationship, as it determines whether economic growth poses a threat to the environment, or rather presents a tool that can be used to combat climate change. We develop a novel semi-parametric panel data methodology that combines country and time fixed effects from the econometrics literature with a so-called feedforward neural network regression component from the machine learning literature. We propose two distinct model specifications: a *static model* and a *dynamic model*. The static model is characterized by assuming the *shape* of the relationship between per capita GDP and per capita CO<sub>2</sub> emissions to be fixed over time and allows only for level-shifts in the relationship over time. This is a common assumption in the literature [Grossman and Krueger 1991] [Holtz-Eakin and Selden] 1995] [Schmalensee, Stoker, and Judson] 1998] [Wagner 2008] [2015]. The novelty of the static model is that a neural network is used to learn the shape of the relationship between per capita GDP and per capita CO<sub>2</sub> emissions. The dynamic model is characterized by being able to learn how the entire shape of relationship between per capita GDP and per capita CO<sub>2</sub> emissions potentially changes over time, which is something not previously considered in the literature. Globally, and for the regions OECD and Asia, we find evidence of an inverse U-shaped relationship between per capita GDP and per capita CO<sub>2</sub> emissions, often referred

to as an environmental Kuznets curve (EKC). Using our dynamic model, we find the global EKC relationship appears stable over time, but the EKC relationship for OECD and Asia does not show up in data until the more recent years. When accounting for emissions transfers across countries, using a consumption-based measure of CO<sub>2</sub> emissions, the EKC relationship for OECD dilutes, suggesting the apparent EKC relationship observed for OECD is driven by emissions exports; for Asia, the EKC relationship becomes even more crisp and exhibits an earlier turning point.

**Chapter 2, *Apocalypse Now? Projecting CO<sub>2</sub> Emissions with Neural Networks*** (joint work with Professor Eric Hillebrand and Associate Professor Mikkel Bennedsen), extends the neural network-based panel data methodology developed in Chapter 1 and proposes a novel *path-dependent model* for constructing CO<sub>2</sub> emissions projections through 2100, using national-level scenarios for per capita GDP from the Shared Socioeconomic Pathways (SSPs; Riahi et al., 2017), also used by the IPCC for the *IPCC Sixth Assessment Report* (IPCC, 2021). Scenario-based CO<sub>2</sub> emissions projections are important, as they help determine whether current climate policies are sufficient for reaching policy goals, such as those put forward by the Paris Agreement (United Nations, 2015), and help to determine the expected costs of emissions reductions. The path-dependent model combines country fixed effects with a long short-term memory (LSTM) recurrent neural network regression component that takes into account time implicitly by building memory and letting model predictions depend on the income path of a country. One of the key features of the model is that it can be readily used out-of-sample, as it avoids explicit dependence on time. Alternative, reduced-form models from the EKC literature rely on the use of time fixed effects, and therefore cannot be readily used out-of-sample, where no time fixed effects are available (Holtz-Eakin and Selden, 1995; Schmalensee et al., 1998; Zhao and Du, 2015). We compare our scenario-based emissions projections to those of so-called integrated assessment models (IAMs), which are large-scale structural models from climate science. For scenarios with low socioeconomic challenges for mitigation SSP1 and SSP4, we find that our emissions projections appear consistent with baseline IAM projections that are meant to describe future developments in absence of new climate policies. For scenarios with medium and high socioeconomic challenges for mitigation SSP2, SSP3, and SSP5, our emissions projections appear the most consistent with IAM mitigation projections that target a forcing level of 6.0 W/m<sup>2</sup> by 2100.

**Chapter 3, *Nowcasting U.S. CO<sub>2</sub> Emissions using Machine Learning***, investigates the use of machine learning methods for nowcasting the yearly growth rate of United States CO<sub>2</sub> emissions. Many important variables used to describe society and the macro economy, such as CO<sub>2</sub> emissions, are sampled at low frequencies and released with a substantial delay. Nowcasting exploits the informational content in a set of explanatory variables that are sampled more frequently than the variable of interest to update an initial forecast before its realized value is released. At the time of writing,

August 2021, official estimates of U.S. CO<sub>2</sub> emissions are available from the inventories of the United Nations Framework Convention on Climate Change<sup>1</sup> only until 2019, highlighting the importance of being able to nowcast CO<sub>2</sub> emissions. Instead of using GDP to explain CO<sub>2</sub> emissions, as we did in Chapter 1 and Chapter 2, I use a high-dimensional panel of macroeconomic variables sampled at mixed frequencies. To handle the problem of mixed frequencies, I propose to represent the data at a common frequency by using the *frequency alignment* transformation from the mixed data sampling regression (MIDAS) literature, whereby which high-frequency variables are transformed into low-frequency vectors through skip-sampling (Ghysels, Kvedaras, and Zemlys, 2016). One of the key features of most machine learning methods, which I exploit in this chapter, is that they solve the so-called “curse of dimensionality” through built-in variable selection mechanisms and feature extraction. In contrast to traditional approaches from the econometrics literature, such as MIDAS models (Ghysels, Santa-Clara, and Valkanov, 2004, 2005, Ghysels, Sinko, and Valkanov, 2007, Andreou, Ghysels, and Kourtellos, 2010), this allows the high-dimensional set of macroeconomic variables to be fed into the machine learning methods without need of initial dimensionality reduction or low-frequency aggregation, and without the need to pre-impose a lot of parametric structure. I find that neural networks, and to a lesser extent tree-based machine learning methods (random forest, bagging, and gradient boosting), are able to utilize the stream of macroeconomic information that becomes available as we progress through the year to produce repeatedly more accurate nowcasts of U.S. CO<sub>2</sub> emissions that generally outperform forecasts from univariate time series models and nowcasts from MIDAS models.

---

<sup>1</sup> Accessible at [di.unfccc.int/time\\_series](https://di.unfccc.int/time_series), last accessed on August 24, 2021.

## References

- Andreou, E., Ghysels, E., Kourtellos, A., 2010. Regression models with mixed sampling frequencies. *Journal of Econometrics* 158 (2), 246–261, specification Analysis in Honor of Phoebus J. Dhrymes.  
URL <https://www.sciencedirect.com/science/article/pii/S0304407610000072>
- Ghysels, E., Kvedaras, V., Zemlys, V., 2016. Mixed frequency data sampling regression models: the R package midasr. *Journal of statistical software* 72 (1), 1–35.
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2004. The MIDAS Touch: Mixed Data Sampling Regression Models. *Cirano working papers*, CIRANO.  
URL <https://EconPapers.repec.org/RePEc:cir:cirwor:2004s-20>
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return trade-off after all. *Journal of Financial Economics* 76 (3), 509–548.  
URL <https://www.sciencedirect.com/science/article/pii/S0304405X05000140>
- Ghysels, E., Sinko, A., Valkanov, R., 2007. MIDAS Regressions: Further Results and New Directions. *Econometric Reviews* 26 (1), 53–90.  
URL <https://doi.org/10.1080/07474930600972467>
- Grossman, G. M., Krueger, A. B., November 1991. Environmental Impacts of a North American Free Trade Agreement. Working Paper 3914, National Bureau of Economic Research.  
URL <http://www.nber.org/papers/w3914>
- Holtz-Eakin, D., Selden, T. M., 1995. Stoking the fires? CO<sub>2</sub> emissions and economic growth. *Journal of Public Economics* 57 (1), 85 – 101.  
URL <http://www.sciencedirect.com/science/article/pii/004727279401449X>
- IPCC, 2014. AR5 Synthesis Report: Climate Change 2014.
- IPCC, 2021. AR6 Climate Change 2021: The Physical Science Basis.
- Riahi et al., 2017. Supplementray Material: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change* 42, 153 – 168.  
URL <http://www.sciencedirect.com/science/article/pii/S0959378016300681>
- Schmalensee, R., Stoker, T., Judson, R., 02 1998. World Carbon Dioxide Emissions: 1950-2050. *The Review of Economics and Statistics* 80, 15–27.
- United Nations, 2015. Paris Agreement. Adopted on 12 Dcember 2015 and entered into force on 4 November 2016.

- Wagner, M., 2008. The carbon Kuznets curve: A cloudy picture emitted by bad econometrics? *Resource and Energy Economics* 30 (3), 388 – 408.  
URL <http://www.sciencedirect.com/science/article/pii/S0928765507000504>
- Wagner, M., 2015. The Environmental Kuznets Curve, Cointegration and Nonlinearity. *Journal of Applied Econometrics* 30 (6), 948–967.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2421>
- Zhao, X., Du, D., 2015. Forecasting carbon dioxide emissions. *Journal of Environmental Management* 160, 39 – 44.  
URL <http://www.sciencedirect.com/science/article/pii/S0301479715300931>



## DANISH SUMMARY

Denne afhandling består af tre selvstændige kapitler, der hvert omhandler brugen af *machine learning* inden for klimaøkonometri og fokuserer særligt på, hvordan idéer og værktøjer fra økonometri og machine learning kan kombineres for at opnå en bedre forståelse for sammenhængen mellem makroøkonomisk aktivitet og CO<sub>2</sub>-udledning. Ifølge FN's klimapanel udgør CO<sub>2</sub>-udledning den primære drivkraft bag klimaforandring og er især drevet af økonomisk- og befolkningsvækst [IPCC] [2014], hvilket understreger vigtigheden af en solid forståelse for sammenhængen mellem makroøkonomisk aktivitet og CO<sub>2</sub>-udledning.

**Kapitel 1, A Neural Network Approach to the Environmental Kuznets Curve** (fælles med professor Eric Hillebrand og lektor Mikkel Bennedsen), betragter bruttonationalproduktet (BNP) som et aggregeret mål for makroøkonomisk aktivitet og anvender paneldata på landeneveau til at undersøge formen på sammenhængen mellem BNP per indbygger og CO<sub>2</sub>-udledning per indbygger. Det er vigtig for både befolkningen og politikere at have en solid forståelse for formen på denne sammenhæng, da den bestemmer hvorvidt økonomisk vækst udgør en trussel mod miljøet, eller istedet udgør et værktøj, der kan bruges til at bekæmpe klimaforandring. Vi udvikler en ny, semiparametrisk metode til paneldata, der kombinerer brugen af lande- og tidsspecifikke *fixed effects* fra den økonometriske litteratur med en regressionskomponent, der består af et såkaldt *feedforward* neuralt netværk fra machine learning litteraturen. Vi foreslår to forskellige modelspecifikationer: en *statisk model* og en *dynamisk model*. Den statiske model er karakteriseret ved at antage, at *formen* på sammenhængen mellem BNP per indbygger og CO<sub>2</sub>-udledning per indbygger er fast over tid og tillader kun parallelforskydninger af sammenhængen over tid. Denne antagelse er typisk i litteraturen [Grossman og Krueger] [1991] [Holtz-Eakin og Selden] [1995] [Schmalensee et al.] [1998] [Wagner] [2008] [2015]. Det nye aspekt ved den statiske model er, at den bruger et neutralt netværk til at lære formen på sammenhængen mellem BNP per indbygger og CO<sub>2</sub>-udledning per indbygger. Den dynamiske model er karakteriseret ved at være i stand til at lære, hvordan hele formen på sammenhængen mellem BNP per indbygger og CO<sub>2</sub>-udledning per indbygger potentiellemændres over tid, hvilket er noget, der ikke tidligere er blevet undersøgt i litteraturen. For hele verden samlet samt for områderne OECD og Asien finder vi evidens for en invers U-form i sammenhængen mellem BNP per indbygger og CO<sub>2</sub>-udledning per indbygger, hvilket ofte

refereres til som en *environmental Kuznets curve* (EKC). Ved at bruge vores dynamiske model finder vi, at EKC-sammenhængen for hele verden ser ud til at være stabil over tid, men at EKC-sammenhængen for OECD og Asien ikke viser sig i data før de senere år. Når vi tager højde for flytning af CO<sub>2</sub>-udledning på tværs af lande ved at bruge et forbrugsbaseret mål for CO<sub>2</sub>-udledning finder vi, at EKC-sammenhængen for OECD udviskes, hvilket antyder, at EKC-sammenhængen for OECD er drevet af eksport af CO<sub>2</sub>-udledning; for Asien finder vi, at EKC-sammenhængen bliver mere tydelig og udviser et tidligere omdrejningspunkt.

**Kapitel 2, *Apocalypse Now? Projecting CO<sub>2</sub> Emissions with Neural Networks*** (fælles med professor Eric Hillebrand og lektor Mikkel Bennedsen), udvider metoden til paneldata baseret på neurale netværk som vi udviklede i Kapitel 1 og foreslår en *stiafhængig model* til at projicere CO<sub>2</sub>-udledning frem til 2100 ved brug af scenarier på landeniveau for BNP per indbygger fra de såkaldte *Shared Socioeconomic Pathways* (SSPs; Riahi et al. 2017), der også anvendes af FN's klimapanel i deres seneste rapport *IPCC Sixth Assessment Report* (IPCC 2021). Scenariebaserede projektioner for CO<sub>2</sub>-udledning er vigtige, da de kan hjælpe med at belyse, hvorvidt den anvendte klimapolitik er tilstrækkelig til at nå nuværende klimamål, der blev lavet i Parisaftalen (United Nations 2015), samt hjælpe med at bestemme omkostningerne ved at reducere udledningen af CO<sub>2</sub>. Den stiafhængige model kombinerer landespecifikke fixed effects med en regressionskomponent, der består af et såkaldt *long short-term memory (LSTM) recurrent* neutralt netværk, der implicit tager højde for tid ved at opbygge hukommelse og lade modellens prediktioner være afhængige af et lands indkomststi. En af de vigtigste egenskaber ved denne model er, at den umiddelbart kan anvendes på data uden for den stikprøve, der er anvendt til at estimere modellen, da den undgår eksplizit at afhænge af tid. Alternative modeller i reduceret form fra EKC-litteraturen afhænger af tidsspecifikke fixed effects og kan derfor ikke umiddelbart anvendes uden for den stikprøve, der er anvendt til at estimere modellen, da her ikke er tidsspecifikke fixed effects tilgængelige (Holtz-Eakin og Selden 1995; Schmalensee et al. 1998; Zhao og Du 2015). Vi sammenligner vores scenariebaserede projektioner med projektioner fra såkaldte *integrated assessment models* (IAMs), hvilket er strukturelle modeller fra klimavidenskaben. For scenarier med lave socioøkonomiske udfordringer SSP1 og SSP4 finder vi, at vores projektioner er i overensstemmelse med referenceprojektioner fra IAMs, der er tænkt til at beskrive en fremtidig udvikling i fravær af ny klimapolitik. For scenarier med mellemstore- og høje socioøkonomiske udfordringer SSP2, SSP3 og SSP5 finder vi, at vores projektioner er mest i overensstemmelse med projektioner fra IAMs, der igennem klimapolitik er rettet mod en strålingspåvirkning på 6.0 W/m<sup>2</sup> for 2100.

**Kapitel 3, *Nowcasting U.S. CO<sub>2</sub> Emissions using Machine Learning*** undersøger brugen af machine learning metoder til at *nowcaste* den årlige vækstrate i den amerikanske CO<sub>2</sub>-udledning. Mange vigtige samfundsmaessige- og makroøkonomiske variable, såsom CO<sub>2</sub>-udledning, er observeret på lav frekvens og ofte oftentliggjort

med en væsentlig forsinkelse. Nowcasting udnytter informationen i et sæt af forklarende variable, der er observeret på højere frekvens end variablen af interesse, til at opdatere et oprindeligt *forecast* af variablen af interesse før dens realiserede værdi er offentliggjort. I skrivende stund, august 2021, er officielle estimerater for den amerikanske CO<sub>2</sub>-udledning kun tilgængelig via *United Nations Framework Convention on Climate Change*<sup>1</sup> frem til år 2019, hvilket understreger vigtigheden af at være i stand til at nowcaste CO<sub>2</sub>-udledning. I stedet for at bruge BNP til at forklare CO<sub>2</sub>-udledning, som vi gjorde i Kapitel 1 og Kapitel 2, så bruger jeg et højdimensionelt panel af makroøkonomiske variable observeret på forskellig frekvens. For at håndtere problemet omkring variable observeret på forskellig frekvens foreslår jeg at repræsentere alt data på samme frekvens ved at anvende *frequency alignment* transformationen fra *mixed data sampling regression (MIDAS)* litteraturen, hvorigenom variable observeret på højere frekvens end variablen af interesse transformeres til vektorer observeret på samme frekvens som variablen af interesse ved brug af *skip-sampling* (Ghysels et al. 2016). En af de vigtigste egenskaber ved de fleste machine learning metoder som jeg udnytter i dette kapitel er deres evne til at løse den såkaldte "curse of dimensionality" gennem indbyggede mekanismer til variabelselektion og *feature extraction*. I modsætning til traditionelle metoder fra den økonometriske litteratur, såsom MIDAS modeller (Ghysels et al. 2004, 2005, 2007, Andreou et al. 2010), tillader machine learning metoder det højdimensionelle sæt af makroøkonomiske variable at blive anvendt uden først at være nødt til at reducere antallet af dimensioner eller aggregere til en lavere frekvens, og uden at være nød til at pålægge en masse parametisk struktur. Jeg finder, at neurale netværk, og i mindre grad træbaserede machine learning metoder (*random forest*, *bagging*, og *gradient boosting*), er i stand til at anvende den strøm af makroøkonomisk information, der bliver tilgængelig i løbet af året, til i voksende grad at producere mere præcise nowcasts af den amerikanske CO<sub>2</sub>-udledning, der generelt er mere præcise end forecasts fra univariate tidsseriemodeller og nowcasts fra MIDAS-modeller.

---

<sup>1</sup>Tilgængelig via [di.unfccc.int/time\\_series](http://di.unfccc.int/time_series), sidst besøgt den 24. august 2021.

## Litteratur

- Andreou, E., Ghysels, E., Kourtellos, A., 2010. Regression models with mixed sampling frequencies. *Journal of Econometrics* 158 (2), 246–261, specification Analysis in Honor of Phoebus J. Dhrymes.  
URL <https://www.sciencedirect.com/science/article/pii/S0304407610000072>
- Ghysels, E., Kvedaras, V., Zemlys, V., 2016. Mixed frequency data sampling regression models: the R package midasr. *Journal of statistical software* 72 (1), 1–35.
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2004. The MIDAS Touch: Mixed Data Sampling Regression Models. Cirano working papers, CIRANO.  
URL <https://EconPapers.repec.org/RePEc:cir:cirwor:2004s-20>
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return trade-off after all. *Journal of Financial Economics* 76 (3), 509–548.  
URL <https://www.sciencedirect.com/science/article/pii/S0304405X05000140>
- Ghysels, E., Sinko, A., Valkanov, R., 2007. MIDAS Regressions: Further Results and New Directions. *Econometric Reviews* 26 (1), 53–90.  
URL <https://doi.org/10.1080/07474930600972467>
- Grossman, G. M., Krueger, A. B., November 1991. Environmental Impacts of a North American Free Trade Agreement. Working Paper 3914, National Bureau of Economic Research.  
URL <http://www.nber.org/papers/w3914>
- Holtz-Eakin, D., Selden, T. M., 1995. Stoking the fires? CO<sub>2</sub> emissions and economic growth. *Journal of Public Economics* 57 (1), 85 – 101.  
URL <http://www.sciencedirect.com/science/article/pii/004727279401449X>
- IPCC, 2021. AR6 Climate Change 2021: The Physical Science Basis.
- Riahi et al., 2017. Supplementray Material: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change* 42, 153 – 168.  
URL <http://www.sciencedirect.com/science/article/pii/S0959378016300681>
- Schmalensee, R., Stoker, T., Judson, R., 02 1998. World Carbon Dioxide Emissions: 1950-2050. *The Review of Economics and Statistics* 80, 15–27.
- United Nations, 2015. Paris Agreement. Adopted on 12 Dcember 2015 and entered into force on 4 November 2016.
- Wagner, M., 2008. The carbon Kuznets curve: A cloudy picture emitted by bad econometrics? *Resource and Energy Economics* 30 (3), 388 – 408.  
URL <http://www.sciencedirect.com/science/article/pii/S0928765507000504>

- Wagner, M., 2015. The Environmental Kuznets Curve, Cointegration and Nonlinearity. *Journal of Applied Econometrics* 30 (6), 948–967.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2421>
- Zhao, X., Du, D., 2015. Forecasting carbon dioxide emissions. *Journal of Environmental Management* 160, 39 – 44.  
URL <http://www.sciencedirect.com/science/article/pii/S0301479715300931>



# 1

C H A P T E R

## A NEURAL NETWORK APPROACH TO THE ENVIRONMENTAL KUZNETS CURVE

**Mikkel Bennedsen**

*Aarhus University and CReATES*

**Eric Hillebrand**

*Aarhus University and CReATES*

**Sebastian Jensen**

*Aarhus University and CReATES*

### Abstract

We investigate the relationship between per capita gross domestic product and per capita carbon dioxide emissions using national-level panel data for the period 1960–2018. We propose a novel semiparametric panel data methodology that combines country and time fixed effects with a nonparametric neural network regression component. Globally and for the regions OECD and Asia, we find evidence of an inverse U-shaped relationship, often referred to as an environmental Kuznets curve (EKC). For OECD, the EKC-shape dilutes when using consumption-based emissions data, suggesting the EKC-shape observed for OECD is driven by emissions exports. For Asia, the EKC-shape becomes even more clear when using consumption-based emissions data and exhibits an earlier turning point.

## 1.1 Introduction

Since the pre-industrial era, anthropogenic activity has driven atmospheric concentrations of greenhouse gases to reach levels unprecedented at least in the last 800,000 years [IPCC 2014], and effects have been detected throughout the climate system. Carbon dioxide ( $\text{CO}_2$ ) is the most important greenhouse gas and the key driver of climate change. It is estimated that  $\text{CO}_2$  is responsible for around 66% of total human contribution to temperature changes since 1750<sup>1</sup>. The observed increase in emissions is deemed largely driven by income and population growth [IPCC 2014], highlighting the importance of a sound understanding of the relationship between per capita income and emissions. A better understanding of the relationship can provide important information to the public and policy makers alike, as it determines whether economic growth poses a threat to the environment, or rather presents a tool that can be used to combat climate change.

The aim of this paper is to investigate the shape of the income-emissions relationship, and assess whether it varies across regions of the world or changes over time. We are particularly interested in whether we find evidence of an inverse U-shaped relationship, often referred to as an environmental Kuznets curve (EKC). This particular shape has played a prominent role in the literature and is surrounded with controversy. Using national-level panel data for the period 1960-2018, we estimate the relationship between per capita gross domestic product (GDP) and per capita  $\text{CO}_2$  emissions. To avoid conclusions being driven by arbitrary *ex ante* restrictions, we want the employed panel data model to be as agnostic as possible. We also want the employed panel data model to be able to simultaneously make use of cross-country dependencies and allow for cross-country heterogeneity in the shape of the income-emissions relationship. We propose a novel semiparametric panel data methodology that combines parametric fixed effects with a nonparametric regression component consisting of a feedforward neural network. Some parameters of the neural network component are shared across all countries and used to learn common input transformations while other parameters are specific to regions of homogeneous countries to allow for cross-country heterogeneity in the shape of the income-emissions relationship. Existing approaches to estimation of the income-emissions relationship using panel data often pre-impose a lot of parametric structure [Grossman and Krueger 1991] [Holtz-Eakin and Selden 1995], and even if a more agnostic approach is taken through the use of splines [Schmalensee et al. 1998] [Auffhammer and Steinhäuser 2012] or kernel-based regression [Millimet, List, and Stengos 2003] [Bertinelli and Strobl 2005], they do not allow for parameter-sharing across countries.

We propose a modeling framework that contains two distinct model specifications: a *static model* and a *dynamic model*. The static model specification contains

---

<sup>1</sup>Percentage is calculated based on Table 2 of [NOAA 2020] and reflects concentration-based, anthropogenic radiative forcing for 2019 relative to 1750s. Radiative forcing measures the amount by which the Earth's energy budget is out of balance due to human activities.

both country and time fixed effects in addition to a neural network component that uses income as its only input variable. The model is static in the sense that the shape of the income-emissions relationship is assumed to be fixed over time. However, the level of the income-emissions relationship may change over time through the time fixed effects. Instead of using time fixed effects to only allow for level-shifts over time, the dynamic model uses a time variable as an additional input into the neural network component. By doing so, the dynamic model learns how the income-emissions relationship potentially changes its entire shape over time. We demonstrate, in a Monte Carlo experiment, that our proposed methodology is able to identify various functional forms of different complexity, and we also demonstrate its ability to account for country-specific stochastic trends. In addition, we demonstrate that when the simulated income-emissions relationship is sufficiently varying in time, the static model with time fixed effects is not able to capture this, but the dynamic model is able to appropriately capture time-varying income-emissions relationships. Globally and for the regions OECD and Asia, we find evidence of an EKC-relationship using our proposed methodology on territorial emissions data. The global relationship appears rather stable over time. On the other hand, it seems the EKC-shape does not appear until late in the sample period for OECD and Asia.

We also study the importance of international trade patterns for explaining the observed income-emissions relationships. We use two distinct types of CO<sub>2</sub> emissions estimates: territorial emissions (based on production) and consumption-based emissions. The latter accounts for cross-country emissions transfers through international trade. Comparing the relationships observed for these distinct types of emissions estimates allows us to assess the importance of international trade patterns. Using consumption-based emissions in our analysis instead of territorial emissions, the evidence of an EKC-relationship for OECD dilutes. For Asia, the EKC-relationship becomes even more pronounced and with an earlier turning point. These findings suggest the EKC-relationship observed for OECD using territorial emissions is driven by emissions exports to other countries. The EKC-relationship observed for Asia when using territorial emissions is driven by local consumption and accounting for emissions imports results in an earlier turning point of the EKC.

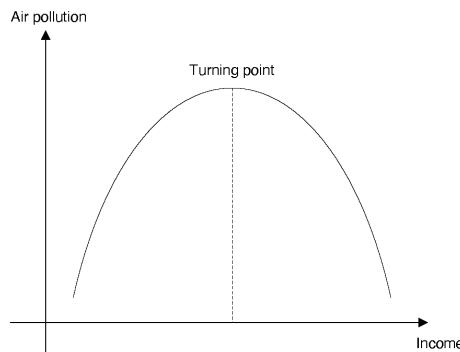
The structure of this paper is as follows. In Section 1.2, we present the fundamental idea of the EKC and discuss econometric issues involved with its estimation. Section 1.3 summarizes the data set used in this paper. In Section 1.4, we present our proposed, neural network-based panel data methodology for analyzing the EKC for CO<sub>2</sub> emissions. Section 1.5 demonstrates the finite sample properties of our proposed methodology through a Monte Carlo experiment, and Section 1.6 presents the empirical analysis. Section 1.7 briefly summarizes and concludes.

## 1.2 The Environmental Kuznets Curve and its Econometric Issues

The literature on reduced-form modeling of the relationship between income and air pollution dates back to the seminal contribution of Grossman and Krueger (1991), who find evidence of an inverse U-shaped relationship between income and various air pollutants. The inverse U-shaped relationship between income and air pollution, sketched in Figure 1.1, is often referred to as the environmental Kuznets curve (EKC). Kuznets (1955) postulates a similar relationship between income and inequality. The EKC is consistent with the idea that, in early stages of economic development, air pollution grows rapidly as production expands (scale effect); however, as income grows, air pollution eventually reaches a turning point after which structural changes (composition effect) and technological improvements (technique effect) lead to a decline. The composition and technique effects may be driven by underlying factors such as increased environmental concerns and environmental policies, which in turn may be driven by even more fundamental factors. Empirically, support of the EKC hypothesis has been mixed. For instance, Selden and Song (1994), Schmalensee et al. (1998), Millimet et al. (2003), Costantini and Martini (2006), and Dutt (2009) find evidence in favor of the EKC hypothesis<sup>2</sup>, while, for example, Shafik (1994), Arrow et al. (1995), Holtz-Eakin and Selden (1995), Wagner (2008), and Stern (2010) find evidence against it. For a recent survey of the EKC literature, see Stern (2017).

Since Grossman and Krueger (1991), the most commonly adopted modeling approach within the EKC literature has been to specify a panel data model consisting of country and time fixed effects in addition to some ad hoc polynomial specification (typically quadratic) of the functional relationship between income and air pollutant. However, this approach likely suffers from a variety of econometric issues (Stern

**Figure 1.1:** Illustrative environmental Kuznets curve




---

<sup>2</sup>We understand evidence of the EKC hypothesis to be evidence of an in-sample turning point in the estimated relationship between income and air pollution.

2004; Müller-Fürstenberger and Wagner 2007; Wagner 2008; Aslanidis 2009). Of these issues, the most critical seem to pertain to functional misspecification, cross-sectional heterogeneity, structural changes, non-identifiable time effects, integrated variables, and omitted variable bias. In this paper, we mainly focus on the issues of functional misspecification, cross-sectional heterogeneity, structural changes, and non-identifiable time effects, but also briefly address the issues of integrated variables and omitted variable bias.

Reliance on ad hoc parametric specifications of the functional relationship between income and air pollutant is problematic, as different parametric specifications can lead to significantly different conclusions, and functional misspecification is likely to occur (Harbaugh, Levinson, and Wilson 2002; Galeotti, Lanza, and Pauli 2006; Tsurumi and Managi 2015). A strand of the EKC literature has therefore focused on semiparametric panel data models that combine use of parametric fixed effects with a nonparametric regression component. This strand of the literature was initiated by Schmalensee et al. (1998), who use splines for the nonparametric component of the model. This strand has focused mainly on spline-based approaches (Harbaugh et al. 2002; Dijkgraaf and Vollebergh 2005; Auffhammer and Steinhauser 2012) and kernel-based approaches (Taskin and Zaim 2000; Millimet et al. 2003; Bertinelli and Strobl 2005; Azomahou, Laisney, and Van 2006). Our work contributes to this discussion by suggesting a semiparametric panel data methodology where the nonparametric component consists of a feedforward neural network, one of the most popular machine learning algorithms. To the best of our knowledge, this is the first time machine learning methods have been applied to study the income-emissions relationship. Although machine learning methods can arguably be statistically opaque, we demonstrate favorable finite sample properties of our proposed methodology in a Monte Carlo experiment.

Assuming the true functional relationship between income and air pollution is of the EKC type, the idea behind relying on national-level panel data is to use information from developing countries to provide information about the positively-sloped part of the EKC, and use information from developed countries to provide information about the negatively-sloped part of the EKC. However, some studies have challenged the assumption that all countries of the world would adhere to a single, global EKC, and advocate allowing for cross-country heterogeneity (de Bruyn, van den Bergh, and Opschoor 1998; List and Gallet 1999; Dijkgraaf and Vollebergh 2005). Some studies also advocate the need to take into account structural changes over time (Halkos and Tsionas 2001; Romero-Ávila 2008; Hendry 2018). We consider two distinct model specifications, distinguished by whether or not they can accommodate changes in the shape of the income-emissions relationship over time. We consider a static model specification, consisting of country and time fixed effects in addition to a neural network component that uses income as its only input variable. The static model imposes the assumption that the shape of the income-emissions relationship is

fixed over time, although it does allow for level-shifts over time through the time fixed effects. We also consider a dynamic model specification. Instead of using time fixed effects to only allow for level-shifts in the income-emission relationship over time, it uses a time variable as an additional input into the neural network component. The dynamic model is able to learn how the income-emissions relationship potentially changes its entire shape over time. In both model specifications, we consider allowing for cross-sectional heterogeneity by letting a subset of parameters vary across regions of the world.

In the EKC literature, country and time fixed effects are usually treated as nuisance parameters, motivating either the random effects or the fixed effects transformation to remove them. However, such an approach might potentially be problematic. The transformations remove only stochastic trends from the data that are common to all countries [Stern and Common 2001]. They cannot remove country-specific stochastic trends. In the presence of country-specific stochastic trends, methods relying on the random effects or the fixed effects transformation will be inconsistent. This situation is likely relevant, since country-specific stochastic trends are often encountered in practice [Perman and Stern 1999]. The importance of being able to properly account for stochastic trends is also emphasized by, for example, [Perman and Stern 2003], [Wagner 2008, 2015], and [Lee and Lee 2009]. Instead of treating country and time fixed effects as nuisance parameters and trying to remove them, we estimate fixed effects explicitly using a dummy variable approach. An extensive Monte Carlo experiment indicates that our proposed methodology is able to account for country-specific stochastic trends.

There is some evidence that the traditional EKC model omits important explanatory variables even after controlling for country and time-specific effects [Stern and Common 2001] [Magnani 2001]. To alleviate the problem of omitted variables, it has been suggested to extend the traditional EKC model with additional explanatory variables reflecting factors such as policies, governance, and institutions [Panayotou 1997] [Torras and Boyce 1998] [Dasgupta, Hamilton, Pandey, and Wheeler 2006]; education, urbanization, and inequality [Hill and Magnani 2002] [Barros, Mendonça, and Nogueira 2002] [Borghesi 2006]; energy prices and fuel mix [Agras and Chapman 1999] [Richmond and Kaufmann 2006], and international trade patterns [Suri and Chapman 1998] [Cole 2003, 2004] [Kearsley and Riddel 2010]. But, no canonical variables have been agreed upon in the literature thus far, and it is outside the scope of this paper to try to identify which variables to include to alleviate problems of omitted variable bias. However, since international trade patterns are often considered one of the most important factors that can potentially explain the EKC [Dinda 2004], we exploit the fact that we have available both territorial and consumption-based CO<sub>2</sub> emissions estimates, and compare the income-emissions relationship obtained using these distinct types of emissions estimates to asses the importance of international trade patterns. We are particularly interested in whether we find evidence of the

much debated pollution haven hypothesis (PHH) (Stern, Common, and Barbier [1996] Dinda [2004]): if changes in the structure of production experienced by developed countries are not met by similar changes in the structure of consumption, the EKC might just reflect displacement of emissions from developed countries with strong environmental regulation to less developed countries with weaker environmental regulation.

### 1.3 Data

To model the relationship between income and, respectively, territorial CO<sub>2</sub> emissions and consumption-based CO<sub>2</sub> emissions, we consider distinct, unbalanced panels of data. The specifications of the panels are reported in Table 1.1.

Territorial CO<sub>2</sub> emissions estimates are from the Global Carbon Project (2019). Estimates include emissions from fossil fuel combustion, oxidation, and cement production, and exclude emissions from bunker fuels, as the latter cannot be allocated unambiguously to particular countries. For the year 2018 (and for some countries 2016-2018), estimates are preliminary and made by the Global Carbon Project (GCP) based on energy statistics published by British Petroleum. For the first sample year 1960, the CO<sub>2</sub> panel covers 86 countries, accounts for 79.0% of the world's total population, and 65.9% of the world's total CO<sub>2</sub> emissions. For the period 1990 onward, it covers more than 160 countries each year and accounts for more than 95.0% of the world's total population and CO<sub>2</sub> emissions.

Use of consumption-based CO<sub>2</sub> emissions in climate analysis has been advocated by Peters and Hertwich (2008a,b). To account for emissions transfers via international trade, consumption-based CO<sub>2</sub> emissions adjust territorial CO<sub>2</sub> emissions by adding emissions embedded in imports and subtracting emissions embedded in exports (Peters, Minx, Weber, and Edenhofer, 2011). Consumption-based CO<sub>2</sub> emissions estimates are also from the Global Carbon Project (2019). Throughout the sample period, which does not begin until 1990, the CO<sub>2</sub><sup>C</sup> panel accounts for more than 90.3% of the world's total population and 93.5% of the world's total CO<sub>2</sub> emissions.

Data on population<sup>3</sup> and GDP are from the World Development Indicators database of the World Bank.<sup>4</sup> The GDP series that we consider is constructed from the separate series GDP<sup>5</sup> (current local currency units), the GDP deflator<sup>6</sup> and a purchasing power parity (PPP) conversion factor<sup>7</sup>.

Besides estimating a global income-emissions relationship, we consider allowing for region-specific relationships. It would be possible to cluster countries into regions

<sup>3</sup>The series "SP.POP.TOTL" was downloaded on December 12, 2019.

<sup>4</sup>Accessible at [databank.worldbank.org/source/world-development-indicators](http://databank.worldbank.org/source/world-development-indicators), last accessed on December 12, 2019.

<sup>5</sup>The series "NY.GDP.MKTP.CN" was downloaded on December 12, 2019.

<sup>6</sup>The series "NY.GDP.DEFL.ZS" was downloaded on December 12, 2019.

<sup>7</sup>The series "PA.NUS.PPP" was downloaded on December 12, 2019.

**Table 1.1:** Panel specifications

CO <sub>2</sub> panel	CO <sub>2</sub> <sup>C</sup> panel	CO <sub>2</sub> * panel
• CO <sub>2</sub> emissions, Mt CO <sub>2</sub>	• CO <sub>2</sub> <sup>C</sup> emissions, Mt CO <sub>2</sub>	• CO <sub>2</sub> emissions, Mt CO <sub>2</sub>
• GDP, billion 2005 USD (PPP)	• GDP, billion 2005 USD (PPP)	• GDP, billion 2005 USD (PPP)
• Population, millions	• Population, millions	• Population, millions
• 1960-2018	• 1990-2017	• 1990-2017
• 186 (81) countries	• 117 (106) countries	• 117 (106) countries
• 8,641 observations	• 3,232 observations	• 3,232 observations

Note: The CO<sub>2</sub> panel contains data on territorial CO<sub>2</sub> emissions estimates (CO<sub>2</sub> emissions) and is used in the territorial emissions analysis of Section 1.6.1, the CO<sub>2</sub><sup>C</sup> panel contains data on consumption-based CO<sub>2</sub> emissions estimates (CO<sub>2</sub><sup>C</sup> emissions) and is used in the consumption-based emissions analysis of Section 1.6.2 and the CO<sub>2</sub>\* panel is based on data from the CO<sub>2</sub> panel, restricted to the same sample period as the CO<sub>2</sub><sup>C</sup> panel, and is used as reference in the consumption-based emissions analysis of Section 1.6.2. CO<sub>2</sub> emissions is measured in megatonnes (Mt; i.e. 10<sup>6</sup> tonnes). Each panel also contains data on GDP measured in United States dollars (USD), adjusted using purchasing power parities (PPP), and population sizes measured in millions. Row four of the table reports sample lengths for each panel. Row five reports the number of countries in each panel, and numbers inside parentheses denote the number of countries with complete data, i.e. countries without “missing data” entries. The final row reports the total number of observations in each panel.

in an initial step using a data-driven approach. However, to facilitate integration of our results within the climate change research community, we specify regions in accordance with the macro-regions defined for the Shared Socioeconomic Pathways (SSPs) and related integrated assessment scenarios of the future (Riahi et al., 2017), used for example by the Coupled Model Intercomparison Project Phase 6 (CMIP6) of the World Climate Research Programme. The regions are defined as follows:

- OECD: 43 OECD90 and EU member states and candidates.
- REF: 13 reforming economies of Eastern Europe and the former Soviet Union.
- Asia: 35 Asian countries excluding the Middle East, Japan and former Soviet Union states.
- MAF: 64 countries of the Middle East and Africa.
- LAM: 34 countries of Latin America and the Caribbean.

Table A.1 of the appendix maps each country to one of the five regions. Tables A.2 and A.3 of the appendix report descriptive statistics for the five regions. Table A.2 contains mean values for the data set. By some margin, OECD has been the richest region throughout the sample period. By contrast, Asia and MAF have been the poorest. There seems to be a tendency that per capita emissions are the highest for the most rich regions. Table A.3 contains standard deviations for the data set. It seems OECD is perhaps the most heterogeneous region, followed by MAF. These two regions are also composed of the highest number of individual countries.

Table A.2 also indicates whether a region is net exporter or net importer of CO<sub>2</sub> emissions: if the level of consumption-based emissions is higher than for territorial emissions, it suggests the region is a net exporter of CO<sub>2</sub> emissions (and vice versa). By relying only on observations for which we have available both consumption-based and territorial emissions (compare means for CO<sub>2</sub>\* and for CO<sub>2</sub><sup>C</sup> in Table A.2), we aim to control for compositional differences in the CO<sub>2</sub> panel and the CO<sub>2</sub><sup>C</sup> panel. Throughout the sample period, OECD is a net exporter of CO<sub>2</sub> emissions, while the regions REF, MAF, and Asia are net importers of CO<sub>2</sub> emissions. LAM goes from having a very balanced import and export of CO<sub>2</sub> emissions in the beginning of the sample period, to being a net importer of CO<sub>2</sub> emissions, to finally being a net exporter of CO<sub>2</sub> emissions. In the empirical analysis in Section 1.6 we discuss whether being a net exporter or net importer of CO<sub>2</sub> emissions can help explain some of the observed income-emissions relationships.

## 1.4 Methodology

The reduced-form relationship between income and greenhouse gas emissions can be mathematically represented in a panel data framework as

$$y_{it} = f(x_{it}, i, t) + u_{it}, \quad i = 1, \dots, N_t, \quad t = 1, \dots, T, \quad (1.1)$$

where  $y$  is a measure of emissions,  $x$  is a measure of income, and  $u$  is an error term. Here,  $i$  indexes countries (or some other cross-sectional units) and  $t$  indexes time periods. Equation (1.1) allows for an unbalanced panel of data by letting  $N_t \leq N$  denote the number of countries observed in time period  $t$  of the total number of countries  $N$ . The function  $f$  characterizes the functional relationship between income and emissions and is the main object of interest to us. Without imposing further restrictions, the function  $f$  cannot be identified from equation (1.1) as only one observation pair  $(y_{it}, x_{it})$  is available for each combination of country and time period  $(i, t)$ . Unfortunately, it is not clear which identifying restrictions are appropriate. As noted by Vollebergh, Melenberg, and Dijkgraaf (2009), different ex-ante restrictions might be driving the non-robustness that plagues the literature on reduced-form estimation of the income-emissions relationship using panel data, causing mixed empirical evidence of the EKC hypothesis. As discussed by Vollebergh et al. (2009), a

particularly important but often overlooked issue relates to the fact that income and emissions are both time dependent. As both variables depend on time, separation of the effect of income from the effect of time hinges crucially on the restrictions used for identification. On the one hand, if one allows fully flexible time effects that are also cross-sectional specific, all variation in the data will be explained by these control variables. On the other hand, if one overly restricts time and cross-sectional effects, too much of the explained variation will be attributed to the income effect.

In the EKC literature, it is standard to identify the function  $f$  in (1.1) by imposing the restriction that it is quadratic and common to all cross-sectional units and time periods,  $f(x_{it}, i, t) = f(x_{it})$ , and that  $u$  is a composite error term consisting of additively separable cross-section and time effects in addition to an idiosyncratic and purely stochastic effect. In this case, equation (1.1) reduces to a model of the form

$$y_{it} = \alpha_i + \beta_t + \delta_1 x_{it} + \delta_2 x_{it}^2 + v_{it}, \quad (\text{Traditional EKC model})$$

where  $\alpha$  and  $\beta$  are treated either as random or fixed effects, and  $v$  is a remainder stochastic error term. If  $\alpha$  and  $\beta$  are correlated with income, a random effects model cannot be estimated consistently (Mundlak [1978]). Therefore, a fixed effects treatment is often preferred.

The time and cross-section effects are supposed to control for omitted variables that are not endogenous consequences of income changes. The country fixed effects  $\alpha$  are supposed to capture exogenous and persistent cross-sectional differences in features such as fossil fuel availabilities and prices, output mixes, regulatory structures, policies, and tastes. The time fixed effects  $\beta$  are supposed to capture time-varying omitted variables and shocks that are common to all countries (Stern [2017]). They capture effects on emissions over time in absence of changes in income, and we mainly interpret them as effects of common technology shocks that are not captured by income changes. If omitted variables are correlated with income and not properly controlled for by the country and time fixed effects, then the slope coefficients  $\delta_1$  and  $\delta_2$  will capture both direct and indirect effects of changes in income on emissions. In a typical EKC study,  $y$  is log-transformed per capita emissions of a greenhouse gas like CO<sub>2</sub> and  $x$  is log-transformed per capita GDP. The focus on per capita quantities reflects the hypothesis that population sizes do not affect average behavior. The log-linear specification is typically preferred over a linear specification as multiplicative cross-sectional and time specific effects are deemed more plausible than additive effects given the heterogeneity of cross-sectional units in a typical study (Schmalensee et al. [1998]). Linear specifications have also been considered in the literature, but no large differences between the two specifications are typically observed (Holtz-Eakin and Selden [1995]).

Although the traditional EKC model has been extensively applied within the literature, the appropriateness of it has been heavily debated (Stern [2004] [2017], Müller-Fürstenberger and Wagner [2007], Wagner [2008]). Rather than pre-imposing the restrictions that  $f$  is necessarily quadratic and common to all cross-sectional

units and time periods, and that  $u$  is a composite error term consisting of additively separable cross-section, time, and idiosyncratic effects, we consider a number of different identifying restrictions and their implications for the estimated income-emissions relationship. Throughout, we drop the parametric restriction that  $f$  is necessarily quadratic. Our goal is to represent  $f$  using a feedforward neural network, imposing as few parametric restrictions as possible on equation 1.1.

In what follows, we present our proposed methodology. In Section 1.4.1 we propose a static neural network model of the form

$$y_{it} = \alpha_i + \beta_t(r) + f^{\text{NN}}(x_{it}, r) + \nu_{it}, \quad (\text{Static neural network model})$$

where the superscript on  $f$  highlights that we use a neural network to model  $f$  non-parametrically, and  $r$  is a regional indicator. A function  $\tau: \{1, \dots, N\} \rightarrow \{1, \dots, R\}$ ,  $i \mapsto r$ ,  $1 \leq R \leq N$ , is initially used to map each country to a macro-region within which countries are assumed to admit the same time effects and functional relationship  $f$  (up to country-specific intercept shifts). This model specification retains the restriction that  $u$  is a composite error term consisting of additively separable country and time effects in addition to an idiosyncratic effect, but drops the restriction that  $f$  is necessarily quadratic and common to all countries. The term *static* used in the name of this model specification refers to the fact that the unknown function of interest  $f$  is assumed to be time-invariant in line with the traditional approach of the literature. In Section 1.4.2 we propose a dynamic neural network model of the form

$$y_{it} = \alpha_i + f^{\text{NN}}(x_{it}, t, r) + \nu_{it}, \quad (\text{Dynamic neural network model})$$

which drops the restriction of additively separable time effects. Instead,  $f$  is allowed to depend explicitly on  $t$ . In this way, the entire shape of  $f$  is allowed to change over time. This is the sense in which this model specification is *dynamic*. By dropping the restriction of separable income and time effects altogether, and leaving it up to the neural network to learn how income and time together affect emissions, this model specification presents a fundamentally new approach to dealing with the issue of non-identifiable separate income and time effects in a reduced form panel model of the income-emissions relationship.

### 1.4.1 Static neural network model

The static neural network model imposes the restrictions that the shape of the functional relationship between income and emissions  $f$  is time invariant, and that the error term  $u$  is a composite error term consisting of additively separable country and time fixed effects in addition to an idiosyncratic and purely stochastic effect. However, we impose no parametric restrictions on the functional form of  $f$ . Instead, we learn  $f$  directly from data using a feedforward neural network. We allow for cross-sectional heterogeneity by initially mapping each country to a region of comparable countries

according to  $\tau : \{1, \dots, N\} \rightarrow \{1, \dots, R\}; i \mapsto r, 1 \leq R \leq N$ , then allow a subset of model parameters to vary across regions. The model can be mathematically represented as

$$y_{it} = \alpha_i + \beta_t(r) + \phi(r)^\top z_{it}^{(h)} + \nu_{it}, \quad i = 1, \dots, N_t, t = 1, \dots, T, r = \tau(i), \quad (1.2)$$

$$z_{it}^{(h)} = g(\kappa^{(h)} + \Gamma^{(h)} z_{it}^{(h-1)}), \quad (1.3)$$

⋮

$$z_{it}^{(2)} = g(\kappa^{(2)} + \Gamma^{(2)} z_{it}^{(1)}), \quad (1.4)$$

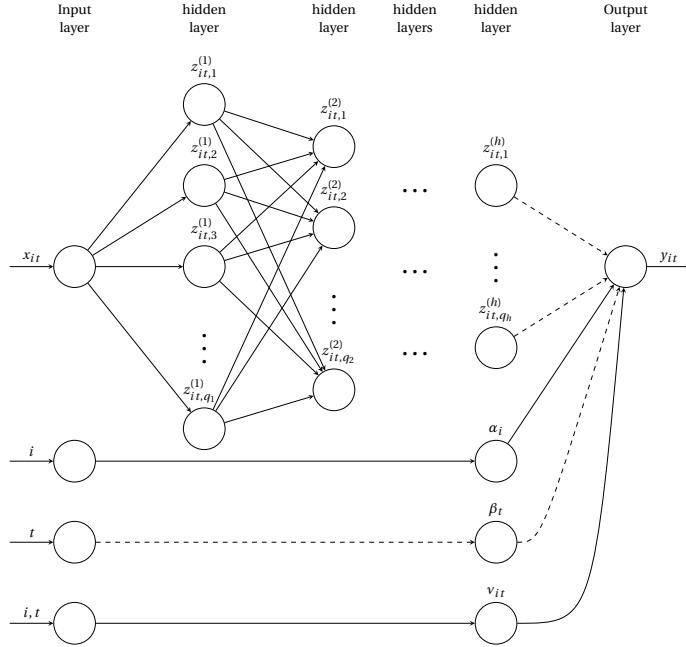
$$z_{it}^{(1)} = g(\kappa^{(1)} + \Gamma^{(1)} x_{it}), \quad (1.5)$$

where  $x_{it}, y_{it} \in \mathbb{R}$  are the natural logarithm of, respectively, per capita GDP and per capita emissions of country  $i$  in year  $t$ . The vector  $\phi(r) \in \mathbb{R}^{q_h}$  is a column vector of region-specific slope coefficients, and  $z_{it}^{(\ell)} \in \mathbb{R}^{q_\ell}$ ,  $\ell = 1, \dots, h$ , is a column vector of variables derived by the neural network. Equation (1.2), which we will refer to as the output layer of the model, is a linear regression model in derived variables  $z_{it}^{(h)}$  augmented by country fixed effects  $\alpha_i$  and region-specific time fixed effects  $\beta_t(r)$ . The derived variables constitute nonlinear transformations of the input variable  $x_{it}$ , learned through so-called hidden layers of the neural network given by (1.3)–(1.5). Instead of  $y_{it}$  depending linearly on  $x_{it}$  and  $x_{it}^2$  as in the traditional EKC model, we learn optimal input transformations directly from the data. Note parameters of the hidden layers are common to all countries. The neural network exploits cross-country dependencies by using the entire panel of data to learn optimal input transformations. Nevertheless, we let slope coefficients of the output layer  $\phi(r)$  be region-specific to allow the shape of the income-emissions relationship to vary across regions. Different choices for the number of regions  $R$  provide different bias-variance tradeoffs by relying on different degrees of parameter sharing across countries. At the one extreme,  $R = 1$  results in a global model with one, global income-emissions relationship (shape) and only level shifts across countries (country fixed effects). At the other extreme,  $R = N$  results in a national model where each country has its own income-emissions relationship. In between those extremes,  $1 < R < N$ , we have a regional model with  $R$  region-specific income-emissions relationships. In the empirical analysis of Section 1.6 we consider  $R = 1$ ,  $R = N$ , and  $R = 5$ , where the five regions considered are OECD, REE, Asia, MAF, and LAM, defined in Section 1.3. In equations (1.2) and (1.3),  $h \geq 1$  denotes the number of hidden layers used to learn the derived variables and is said to determine the *depth* of the neural network;  $q_\ell$  denotes dimensionality of the  $\ell$ -th hidden layer and is referred to as *width*. This implies that  $z_{it}^{(\ell)}$  is a  $q_\ell$ -dimensional column vector of derived variables,  $\kappa^{(\ell)}$  is a  $q_\ell$ -dimensional column vector of unknown intercepts<sup>8</sup> to be estimated from data, and  $\Gamma^{(\ell)}$  is a  $q_\ell \times q_{\ell-1}$  coefficient matrix<sup>9</sup> to be estimated from data. We follow the convention to set  $q_0 \equiv 1$ ,

---

<sup>8</sup>Typically referred to as *biases* in the neural network literature.

<sup>9</sup>Typically referred to as *weights* in the neural network literature.

**Figure 1.2:** Static neural network model

Note: The notation  $z_{it,j}^{(\ell)}$  refers to the  $j$ -th element of the vector of derived variables in the  $\ell$ -th layer for country  $i$  at time  $t$ . Dashed edges in the graph represent region-specific connections.

implying that  $\Gamma^{(1)}$  is a  $q_1$ -dimensional column vector. The non-linear function  $g$  is referred to as an *activation function*. The nonlinearity of  $g$  is what allows the network to learn nonlinear relationships. We rely on the Swish activation function of Ramachandran, Zoph, and Le (2017), which is found to outperform the standard rectified linear unit (ReLU) activation function (Glorot, Bordes, and Bengio 2011) on a number of different tasks (Ramachandran et al. 2017). The Swish function is defined as  $g(z) = z(1 + \exp(-z))^{-1}$  for  $z \in \mathbb{R}$ , and can be considered a smoothed version of the ReLU function defined as  $g(z) = \max(z, 0)$ . We prefer the Swish function as it provides a slightly smoother  $f$  in cases where only one input variable is passed through the hidden layers of the neural network.

The choices of overall network depth and width of each hidden layer constitute important architectural considerations with implications for the representation capabilities of the network. It can be shown that even a shallow feedforward network with just one hidden layer can approximate any Borel-measurable function from one finite-dimensional space to another to any desired degree of accuracy provided the network is wide enough (Hornik, Stinchcombe, and White 1989; Cybenko 1989).

Leshno, Lin, Pinkus, and Schocken [1993]. However, adding depth to a network allows it to learn increasingly more abstract representations of the input data, and, typically, less wide networks with fewer parameters overall are required for a given level of accuracy when compared to a shallow network. On the other hand, deeper networks are typically more difficult to optimize than shallow ones, and require more input data to avoid overfitting to in-sample noise [Goodfellow, Bengio, and Courville [2016]]. In Section 1.4.4 we discuss how we decide on an optimal architecture.

The static model is visually represented in Figure 1.2 by means of a directed acyclic graph (DAG). The upper part of the graph illustrates the neural network component. The bottom part of the graph illustrates the parametric component. Edges represent how information travels through the model. The neural network is feedforward as information only travels forward through the network without any feedback loops. Vertices of the input layer reflect information that is presented to the model. Vertices of the hidden layers reflect elements of the sequentially derived  $z$ -variables. Remaining vertices represent elements of the error term and the output.

#### 1.4.2 Dynamic neural network model

For the dynamic neural network model, we drop the restriction of a time-invariant functional relationship  $f$ , and instead let the neural network learn how the functional relationship depends on time. We do so by dropping the assumption of time fixed effects and instead pass a time variable through the hidden layers of the neural network together with income. Again, we allow for cross-sectional heterogeneity by initially mapping each country to a region of comparable countries according to  $\tau : \{1, \dots, N\} \rightarrow \{1, \dots, R\}$ ,  $i \mapsto r$ ,  $1 \leq R \leq N$ , then allow a subset of model parameters to vary across regions. The model can be mathematically represented as

$$y_{it} = \alpha_i + \phi(r)^\top z_{it}^{(h)} + v_{it}, \quad i = 1, \dots, N_t, t = 1, \dots, T, r = \tau(i), \quad (1.6)$$

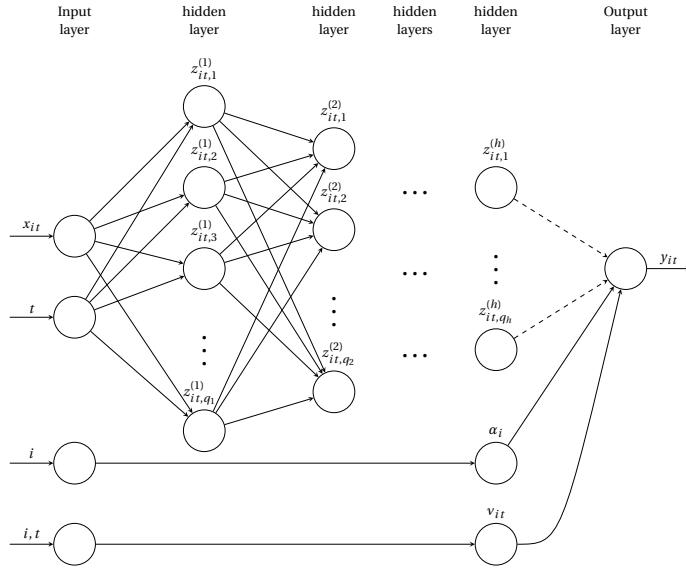
$$z_{it}^{(h)} = g\left(\kappa^{(h)} + \Gamma^{(h)} z_{it}^{(h-1)}\right), \quad (1.7)$$

$\vdots$

$$z_{it}^{(2)} = g\left(\kappa^{(2)} + \Gamma^{(2)} z_{it}^{(1)}\right), \quad (1.8)$$

$$z_{it}^{(1)} = g\left(\kappa^{(1)} + \Gamma^{(1,1)} x_{it} + \Gamma^{(1,2)} t\right), \quad (1.9)$$

with  $x_{it}, y_{it}, z_{it}^{(\ell)}, \phi, \kappa^{(\ell)}, \Gamma^{(\ell)}$  as above, and  $\Gamma^{(1,1)}, \Gamma^{(1,2)}$  being  $q_\ell$ -dimensional column vectors of unknown slope coefficients. Once again, (1.6) is a linear regression model in derived variables  $z^{(h)}$ , learned through hidden layers of a neural network, augmented by country fixed effects  $\alpha_i$ . What is new is that  $z_{it}^{(h)}$  is derived from both income and time. In the static model formulation,  $z_{it}^{(h)}$  is derived from income only. We note that, because there is only one observation pair  $(x_{it}, y_{it})$  available for each country at each point in time, we cannot sensibly consider a national formulation ( $R = N$ )

**Figure 1.3:** Dynamic neural network model

Note: The notation  $z_{it,j}^{(\ell)}$  refers to the  $j$ -th element of the vector of derived variables in the  $\ell$ -th layer for country  $i$  at time  $t$ . Dashed edges in the upper part of the graph represent region-specific connections.

of the dynamic model as this would result in a perfect fit. The DAG associated with the dynamic model is shown in Figure 1.3. By passing a time variable through the hidden layers of the neural network together with income, the network is able to learn how the two variables interact, and hence how the functional relationship between income and emissions potentially changes shape over time.

#### 1.4.3 Estimation

For both the static and dynamic model, we estimate<sup>10</sup> all free parameters simultaneously by minimizing the sum of region-specific mean squared errors:

$$J(\phi, \kappa, \Gamma, \alpha, \beta) = \sum_{r=1}^R \sum_{t=1}^T \sum_{i \in I_r} \frac{1}{n_r} (y_{it} - \hat{y}_{it}(x_{it}))^2, \quad (1.10)$$

where  $n_r$  is the number of observations for region  $r$ , and  $I_r \subseteq \{1, 2, \dots, N\}$  is the set of indices of countries belonging to region  $r$ . Note the model output  $\hat{y}$  is dependent on the estimated parameter vectors  $\hat{\phi}$ ,  $\hat{\kappa}$ ,  $\hat{\Gamma}$ ,  $\hat{\alpha}$ , and possibly  $\hat{\beta}$ , which are understood to contain model parameters of all relevant layers and regions. A possible alternative to

---

<sup>10</sup>Often referred to as *training* in the neural networks literature.

the loss function in (1.10) would be to minimize global mean squared errors where each contribution to the loss function is weighted by  $1/n$  instead of  $1/n_r$ . However, by specifying the loss function as a sum of region-specific mean squared errors, we incorporate a tradeoff between the fit on each region. We ensure the estimated model has an accurate fit for each region separately and that the model is not biased toward regions with more observations than others. Note the weights have no effect on parameter estimates in the global ( $R = 1$ ) formulation of the models.

Before discussing practical details concerning minimization of  $J$ , we briefly elaborate on a subtlety regarding identification of fixed effects. The neural network is so flexible that it automatically adapts its predictions to the level of the dependent variable in the estimation sample. From a purely illustrative point of view, we can think of the neural network as implicitly identifying an overall intercept even if not specified explicitly. This is true even if imposing restrictions such as  $f(0) = 0$ . For this reason, we can only identify country fixed effects for all but one country and time fixed effects for all but one time period. In all applications of this paper, we use the U.S. as a reference when estimating a global model formulation ( $R = 1$ ), then include dummies for the remaining countries. Hence, estimated income-emissions relationship will be normalized for the U.S., and estimated country fixed effects should be interpreted as intercept shifts relative to the U.S. When estimating a regional model formulation ( $R = 5$ ), we must account for the fact that we are in a sense estimating five neural networks simultaneously. Therefore, we must normalize the income-emissions relationships for each region separately. We use the U.S. as reference for OECD, Russia as reference for REF, China as reference for Asia, South Africa as reference for MAF, and Mexico as reference for LAM. When estimating a national formulation ( $R = N$ ) of the static model, we cannot identify any country fixed effects. When estimating the static model, we also use the initial time period  $t = 1$  as reference, then include dummies for remaining time periods  $t = 2, \dots, T$ . Hence, estimated time fixed effects should be interpreted as intercept shifts relative to the initial time period. When estimating a regional formulation of the static model, we allow region-specific time fixed effects. When estimating a national formulation of the static model, we exclude time fixed effects: since there is only one observation pair  $(x_{it}, y_{it})$  available for each country at each point in time, including (country-specific) time fixed effects would result in a perfect fit.

We follow the standard in the neural networks literature and minimize  $J$  using gradient descent. Since the optimization problem is high-dimensional and nonconvex, the loss function most likely features numerous critical points. Whereas most second-order methods would get stuck in any critical point, gradient descent is often able to escape critical points associated with a high loss (Goodfellow et al., 2016). However, naive gradient descent performs poorly whenever the Hessian matrix has a poor condition number. Without modification, gradient descent does not know to prefer directions of parameter space where the slope of the loss function remains

negative for longer. Poor conditioning of the Hessian matrix also makes it difficult to determine a good step size for the gradient descent algorithm.<sup>11</sup> One must balance the goals of avoiding to overshoot the minimum and being able to make significant progress in directions with little curvature. We use the popular Adam (ADaptive Moment estimation) variant of gradient descent. It seeks to mitigate these issues through the use of momentum and a separate learning rate for each parameter that is automatically adapted at each iteration of the algorithm (Kingma and Ba [2014]). The Adam algorithm is illustrated in Figure A.5 of the appendix. For each iteration<sup>12</sup> of the Adam algorithm, we use the entire data set when evaluating the gradient.<sup>13</sup> We use suggested defaults for the hyperparameters of the Adam algorithm (Kingma and Ba [2014]). We find they work well in all applications. We stop the Adam algorithm when we have observed no significant decrease in the loss function over 100 consecutive iterations, using a tolerance level of  $10^{-6}$ , then restore parameter estimates associated with the lowest loss across all iterations (not necessarily the last iteration).

Since Adam is based on local moves, it works well only if initialized within a well-behaved region of parameter space that is connected to a satisfactory solution by a path the algorithm can follow. It is therefore important to use appropriate initial values for  $\alpha$ ,  $\beta$ ,  $\phi$ ,  $\kappa$ , and  $\gamma$ . We initialize all slope coefficients randomly from a truncated normal distribution as suggested by He, Zhang, Ren, and Sun [2015]. This is the suggested way to break symmetry (Goodfellow et al. [2016]), motivate each hidden unit to learn a different function, and avoid problems of vanishing and exploding gradients when using ReLU-type activation functions. We use multiple different initializations to ensure the minimization routine has arrived at a satisfactory minimum. All intercepts (including fixed effects) are initialized from zero as is standard in the neural networks literature (Goodfellow et al. [2016]).

#### 1.4.4 Model selection

The choice of neural network architecture is a standard model selection task: there is a tradeoff between choosing large values for the depth and width parameters to reduce bias, and choosing small values to ensure smoothness. Within the neural networks literature, hypothesis tests, cross-validation, Bayesian regularization, dropout, and information criteria have been proposed for model selection (Anders and Korn [1999]). We use the Bayesian information criterion (BIC). We have also experimented with use of cross-validation, but we found that, in our applications, cross-validation resulted in unreasonably complex network architectures. The Bayesian information criterion (BIC) takes the form

$$\text{BIC} = \log J(\hat{\phi}, \hat{\kappa}, \hat{\Gamma}, \hat{\alpha}, \hat{\beta}) + \frac{m \log n}{n}, \quad (1.11)$$

---

<sup>11</sup>Often referred to as the *learning rate* in the neural networks literature.

<sup>12</sup>Often referred to as an *epoch* in the neural networks literature.

<sup>13</sup>Often referred to as *batch learning* in the neural networks literature.

where  $J(\hat{\phi}, \hat{\kappa}, \hat{\Gamma}, \hat{\alpha}, \hat{\beta})$  is the objective function in (1.10) evaluated in parameters estimates,  $n$  is total number of observations, and  $m$  is total number of model parameters excluding fixed effects.

As discussed by Gu, Kelly, and Xiu (2020), it is not necessary to search over unreasonably many network architectures. We fix a reasonable number of architectures of varying complexity ex ante. We focus on rectangular and pyramid-shaped architectures as proposed by Masters (1993), which are useful for learning gradually more abstract transformations of the input variable and for keeping the number of free parameters at a reasonable level. The full list of candidate architectures is presented in Table A.4 of the appendix together with the number of free model parameters.

## 1.5 Monte Carlo Experiment

In this section, we demonstrate finite sample properties of our proposed methodology in a controlled setup that closely resembles the situation in the empirical analysis of Section 1.6. We demonstrate that our proposed methodology is able to identify various functional forms and account for country-specific stochastic trends. We also demonstrate it is important to use the dynamic model specification in situations where the true relationship of interest is time varying.

### 1.5.1 Static neural network model

To investigate finite sample properties of the static neural network model, we simulate data from the following data-generating process:

$$\begin{aligned} y_{it} &= f(x_{it}, r) + u_{it}, & i = 1, \dots, N_t, t = 1, \dots, T, r = \tau(i), \\ u_{it} &= \alpha_i + \beta(r) \log t + v_{it}, & v_{it} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_v^2), \end{aligned}$$

where  $\tau : \{1, \dots, N\} \rightarrow \{1, \dots, R\}$  is a function mapping each country to one of the  $R$  regions. To obtain as realistic a setup as possible, where  $x$  is characterized by multiple country-specific stochastic trends,  $x$  is taken to be the natural logarithm of per capita GDP from the CO<sub>2</sub> emissions panel discussed in Section 1.3. We consider a global model formulation ( $R = 1$ ) and a regional model formulation with  $R = 5$ . In line with the discussion of Section 1.3, we consider the five regions OECD, REF, Asia, MAF, and LAM. We focus on these model formulations as they are the main focus in Section 1.6. We consider the following different specifications of  $f$ , which one could imagine being confronted with in an empirical EKC application, varying in degree of complexity:

$$\begin{aligned} f(x, r) &= \delta_1(r)x, & (\text{Linear function}) \\ f(x, r) &= \delta_1(r)x + \delta_2(r)x^2, & (\text{Quadratic function}) \\ f(x, r) &= \delta_1(r)x + \delta_2(r)x^2 + \delta_3(r)x^3, & (\text{Cubic function}) \end{aligned}$$

**Table 1.2:** Optimal neural network architectures for the static model in the Monte Carlo experiment

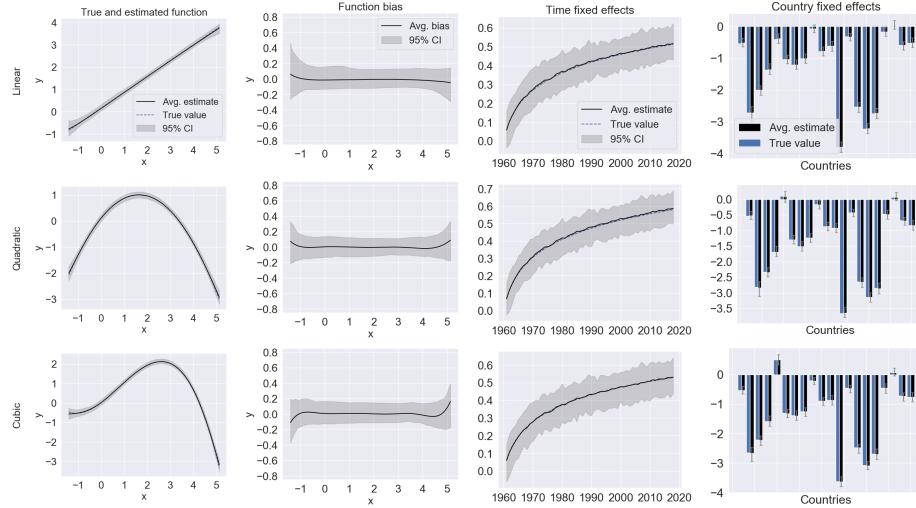
	Function		
	Linear	Quadratic	Cubic
<b>Global model formulation</b>			
Neural network architecture	(2)	(4)	(4)
# parameters (excl. fixed effects)	4	12	12
<b>Regional model formulation</b>			
Neural network architecture	(2)	(4)	(4)
# parameters (excl. fixed effects)	14	28	28

Note: “(a)” indicates a neural network architecture with one hidden layer containing  $a$  units.

where  $\delta_1(r), \delta_2(r), \delta_3(r) \in \mathbb{R}$  are region-specific, constant parameters. To obtain realistic parameter values, we use estimates from simple regressions based on the function under consideration (linear, quadratic or cubic) and the data discussed in Section 1.3. We use the natural logarithm of per capita CO<sub>2</sub> emissions from the CO<sub>2</sub> panel of Section 1.3 as dependent variable and the natural logarithm of per capita GDP from the CO<sub>2</sub> emissions panel and appropriate transformations (square and cube) as input variables. We include dummy variables for each country and a logarithmic time trend. Note that time fixed effects can be backed out from the estimated time trend. For the regional model formulation, we interact input variables and the time trend with a regional indicator to allow region-specific slope coefficients and time effects. Finally, we scale some estimated slope coefficients to achieve a desired amount of variation in  $f$ . We set  $\sigma_v^2 \equiv 0.35$ , which is the empirical standard deviation of the residuals from the simple regression using the natural logarithm of per capita GDP and its square as input variables, without any interactions. For simplicity, we use this value throughout all experiments in the section.

In order to demonstrate the finite sample properties of the proposed estimation procedure, we simulate 100 Monte Carlo samples<sup>14</sup> for each model formulation and the three functions. To keep runtime at a reasonable level, we also construct an initial Monte Carlo sample that is used to determine the optimal network architecture. This architecture, chosen on the initial sample, is then kept for the remaining 100 samples. In this way, the results presented here can be seen as a lower bound on the performance of the estimators, since their properties would likely improve if the optimal architecture was allowed to be chosen separately for each individual Monte Carlo sample. In Table 1.2, we report optimal network architectures for both the global

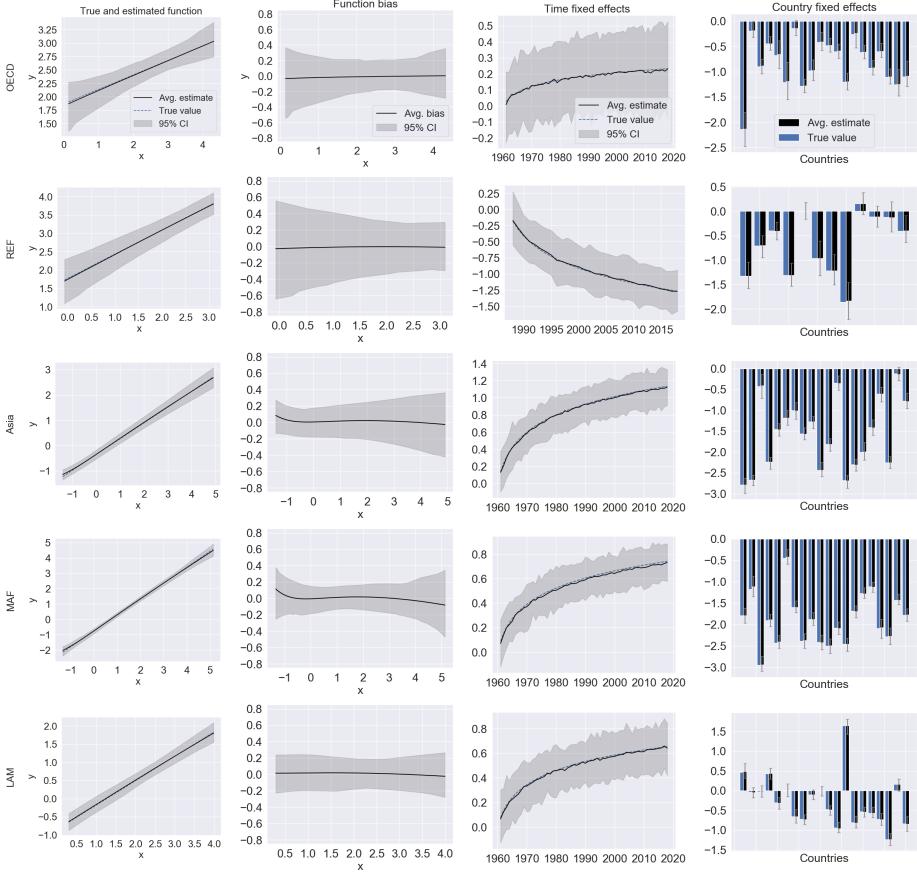
<sup>14</sup>We use only 100 Monte Carlo samples as it is time consuming to estimate the models.

**Figure 1.4:** Monte Carlo results for the global formulation of the static model

Note: In the plots in columns 1 and 3, the true values are covered by the average estimates throughout most of the input region. In the plots in column 4, gray error bars indicate 95% confidence intervals.

and the regional model formulation, determined by minimizing the BIC (1.11) on an initial Monte Carlo sample. As discussed in Section 1.4.4 we search over the set of architectures presented in Table A.4 of the appendix. For both model formulations and all functions, the optimal neural network architecture is a simple one-hidden-layer architecture. For the linear function, two hidden units is optimal for both the global and regional model formulation. However, since the regional model formulation has region-specific slope coefficients in the output layer, it has more free parameters to estimate than the global model formulation for a given network architecture, even without considering the added number of time fixed effects associated with the regional formulation. For the quadratic and cubic function, four hidden units are optimal for both model formulations. Note how it is optimal to slightly increase the complexity of the network architecture, and so the approximation capabilities of the model, as we increase the complexity of the function to approximate. Note also that the models incur a large number of free parameters. For instance, we see from Table 1.2 that the regional formulation of the static model with only a single hidden layer containing four hidden units contains 28 free parameters excluding fixed effects.

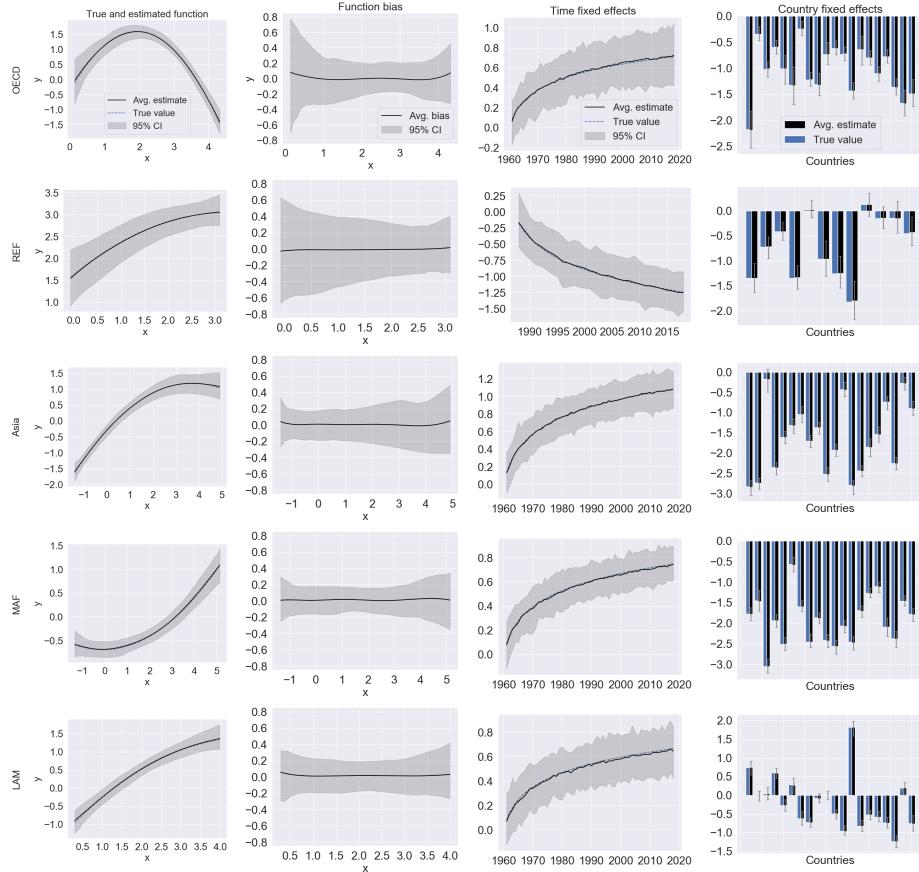
For the global model formulation, average estimates across Monte Carlo samples are reported in Figure 1.4. The first column of Figure 1.4 contains the true functional relationships and average estimate across Monte Carlo samples, plotted over the entire input region used for estimation. We also plot 95% confidence bands obtained from the quantiles of the estimates across Monte Carlo samples. We note that the

**Figure 1.5:** Monte Carlo results for the regional formulation of the static model: linear function

Note: In the plots in columns 1 and 3, the true values are covered by the average estimates throughout most of the input region. In the plots in column 4, gray error bars indicate 95% confidence intervals.

model is able to capture all three functions to a high degree of accuracy with narrow confidence bands. In the second column of Figure 1.4 we plot function bias, i.e. the difference between the curves shown in the first column. We initially note the bias is around zero for all three functions throughout most of the input region. We also note the model displays largest discrepancies toward the boundaries of the input region used for estimation. This is likely related to the sparsity of observations toward the boundaries of the input region used for estimations and the so-called boundary issues well known from the literature on nonparametric regression using kernel-based techniques [Malec and Schienle, 2014]. Although a correction mechanism is outside the scope of this paper, it suggests that one should be careful when interpreting model

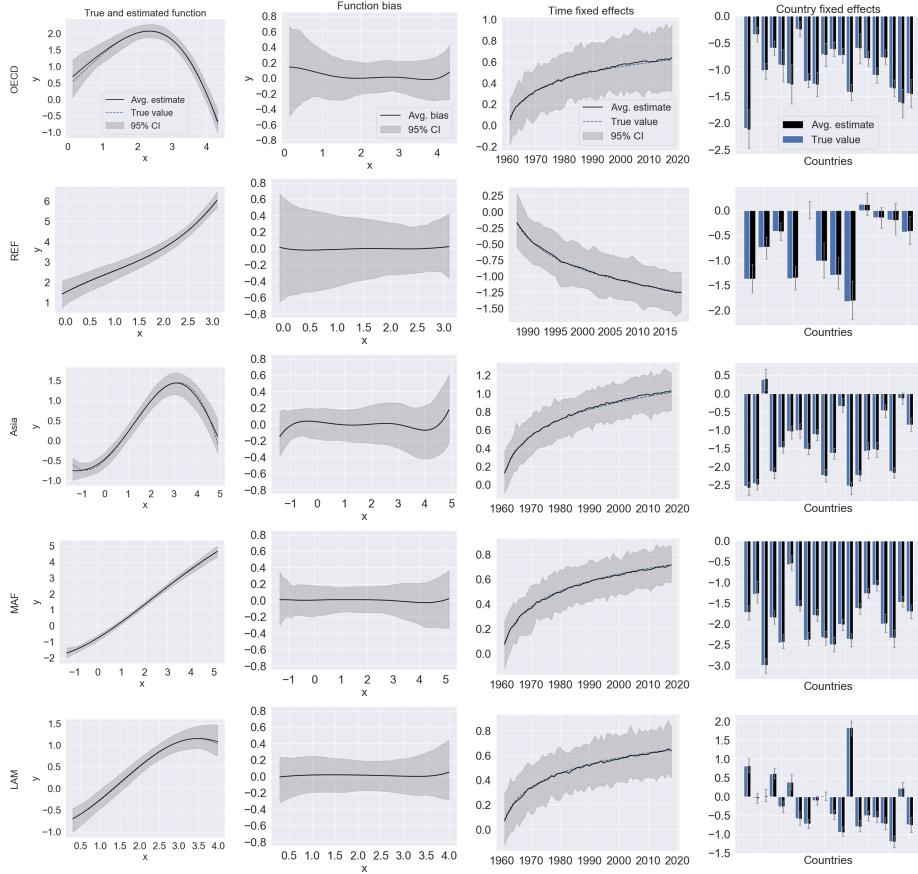
**Figure 1.6:** Monte Carlo results for the regional formulation of the static model: quadratic function



Note: In the plots in columns 1 and 3, the true values are covered by the average estimates throughout most of the input region. In the plots in column 4, gray error bars indicate 95% confidence intervals.

output based on input values close to or beyond the boundary of the input region used for estimation. The third and fourth columns of Figure 1.4 contain true values and average estimates for the time and country fixed effects, respectively. Again, we note the model is able to capture the true values with a high degree of accuracy for all three functions.

Figures 1.5, 1.6, and 1.7 summarize average estimates across Monte Carlo samples for the regional model formulation and the linear, quadratic, and cubic function, respectively. Although the learning task faced by the regional model formulation seems more complex than that faced by the global model formulation, as the regional model formulation is faced with the task of learning region-specific functions and

**Figure 1.7:** Monte Carlo results for the regional formulation of the static model: cubic function

Note: In the plots in columns 1 and 3, the true values are covered by the average estimates throughout most of the input region. In column 4, gray error bars indicate 95% confidence intervals.

time effects, the accuracy of the regional model formulation seems on par with that of the global model formulation in Figure 1.4 for each region separately. It is also encouraging that the model seems to perform about equally well for all regions despite large degree of variation in the number of observations available for each region. For instance, the model has available as little as 373 observation for the region REF and as many as 3008 observations for the region MAF.

### 1.5.2 Dynamic neural network model

To investigate finite sample properties of the dynamic neural network model, we simulate data from the following data-generating process:

$$\begin{aligned} y_{it} &= f(x_{it}, t, r) + u_{it}, \quad i = 1, \dots, N_t, \quad t = 1, \dots, T, \quad r = \tau(i), \\ u_{it} &= \alpha_i + v_{it}, \quad v_{it} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_v^2), \end{aligned}$$

with notation as above. We again consider  $R = 1$  (global model) and  $R = 5$  (regional model). In contrast to the static model, the effect of time is not modeled by a logarithmic time trend. Instead,  $f$  now changes its entire shape over time. Specifically, we consider the following different specifications of  $f$ , which one could imagine facing in an empirical EKC application, varying in degree of complexity:

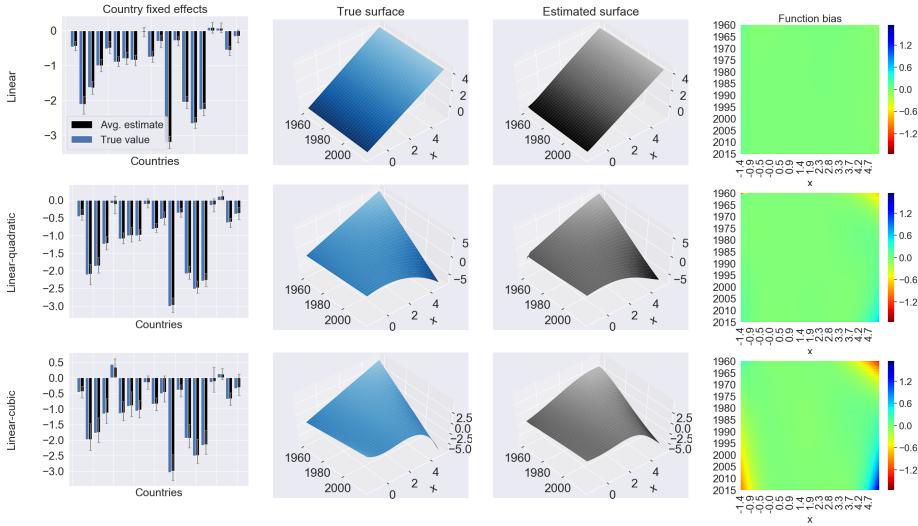
$$\begin{aligned} f(x, t, r) &= \delta_1(r)x, && \text{(Linear function)} \\ f(x, t, r) &= \frac{T-t+1}{T}\delta_1(r)x + \frac{t-1}{T}\delta_2(r)x^2, && \text{(Linear-quadratic function)} \\ f(x, t, r) &= \frac{T-t+1}{T}\delta_1(r)x + \frac{t-1}{T}\delta_2(r)x^2 + \frac{t-1}{T}\delta_3(r)x^3, && \text{(Linear-cubic function)} \end{aligned}$$

where  $\delta_1(r), \delta_2(r), \delta_3(r) \in \mathbb{R}$  are region-specific, constant parameters. To obtain realistic parameter values, we once again resort to estimates from simple regressions based on the function under consideration (linear, linear-quadratic or linear-cubic). Using the linear-cubic function as an example, we run a regression using the natural logarithm of per capita CO<sub>2</sub> emissions from the CO<sub>2</sub> panel of Section 1.3 as dependent variable and the natural logarithm of per capita GDP from the CO<sub>2</sub> emissions panel and its square and cube as input variables. We also include dummy variables

**Table 1.3:** Optimal neural network architectures for the dynamic model in the Monte Carlo experiment

	Function		
	Linear	Linear-quadratic	Linear-cubic
Global model formulation			
Neural network architecture	(2)	(4,4)	(4,2,2)
# parameters (excl. fixed effects)	8	36	30
Regional model formulation			
Neural network architecture	(2)	(4,4)	(4,4)
# parameters (excl. fixed effects)	16	52	52

Note: “ $(a,b,c)$ ” indicates a neural network architecture with three hidden layers containing  $a$  units in the first layer,  $b$  in the second, and  $c$  in the third.

**Figure 1.8:** Monte Carlo results for the global formulation of the dynamic model

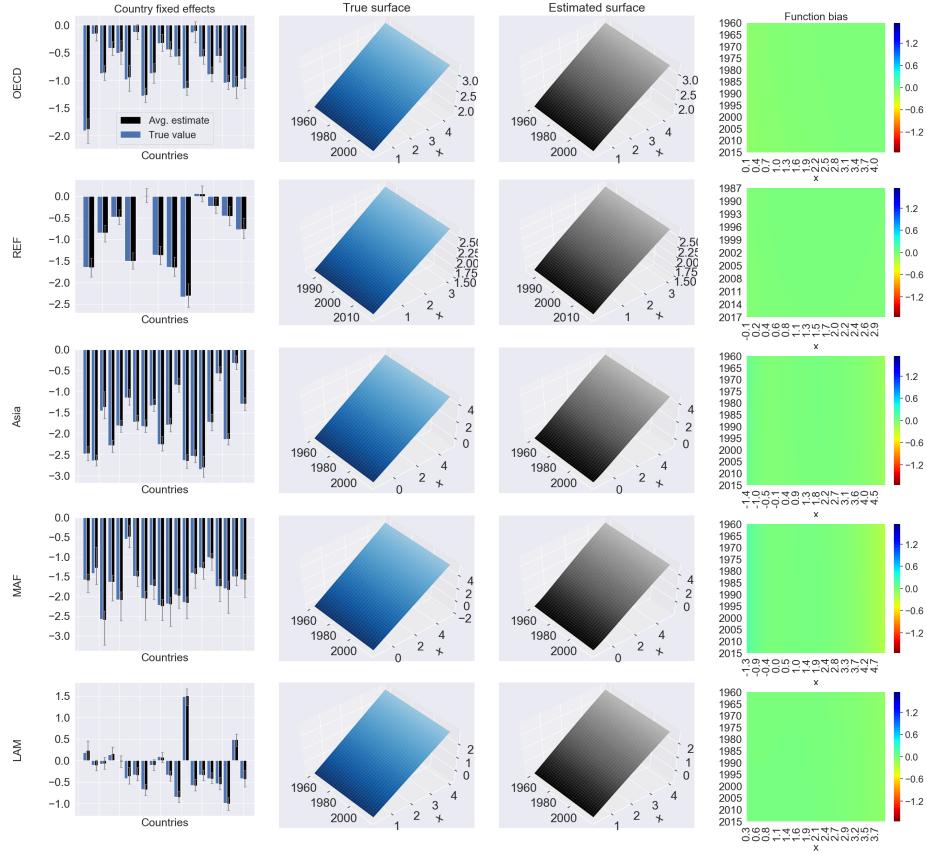
Note: In the country fixed effects plots, gray error bars indicate 95% confidence intervals.

for each country. For the regional model formulation, we interact input variables with a regional indicator to allow region-specific slope coefficients. In this way, we obtain realistic values for the country fixed effects and for  $\delta_1(r)$ ,  $\delta_2(r)$ , and  $\delta_3(r)$ ,  $r = 1, \dots, R$ . Finally, we scale some estimated slope coefficients to achieve a sufficient amount of variation in  $f$ .

Table 1.3 contains optimal network architectures for both the global and regional formulation of the dynamic model obtained by minimizing the BIC (1.11) on an initial Monte Carlo sample as discussed above. For the simple linear function, both model formulations require only a single-hidden-layer architecture with two hidden units, the simplest architectures in the set of candidates. Not counting the number of fixed effects, the dynamic model contains more free parameters than its static counterpart for a given formulation and network architecture because both time and income are passed through the hidden layers of the neural network. The learning task faced by the dynamic model seems in many ways more complex than that of its static counterpart. The dynamic model is required to learn how the functional relationship of interest changes its entire shape over time, whereas the static model allows only for intercept shifts over time. From Table 1.3 we see increased complexity of the learning task prompts an increase in optimal model complexity. In particular, it is optimal to add depth to the network architecture when turning from the linear function to the more complex and time-varying linear-quadratic and linear-cubic functions.

Figure 1.8 demonstrates finite sample properties of the global formulation of the

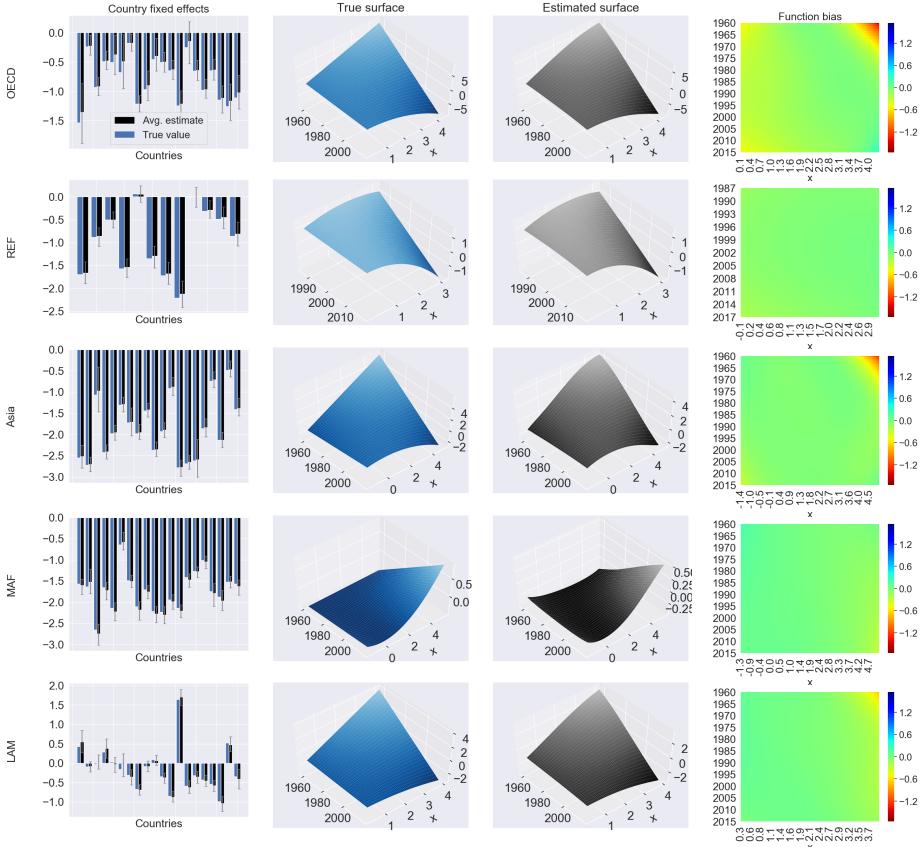
**Figure 1.9:** Monte Carlo results for the regional formulation of the dynamic model: linear function



Note: In the country fixed effects plots, gray error bars indicate 95% confidence intervals.

dynamic model by summarizing average estimates across 100 Monte Carlo samples. The first column of plots shows true and estimated country fixed effects together with 95% confidence intervals obtained from the quantiles of the estimates across Monte Carlo samples. For all three functions, the model is able to accurately estimate the country fixed effects. The second and third columns of plots show, respectively, true and estimated surfaces, plotted over the entire input region used for estimation. The fourth column of plots displays heat maps of function bias obtained by subtracting the surfaces in the second and third columns. Initially, we note that the bias is around zero throughout most of the input region used for estimation. However, similar to the static model, it is more difficult for the model to accurately estimate the true function toward the boundaries of the input region used for estimation. Indeed, worst performance of the model occurs toward the four corners of the input region used

**Figure 1.10:** Monte Carlo results for the regional formulation of the dynamic model: linear-quadratic function

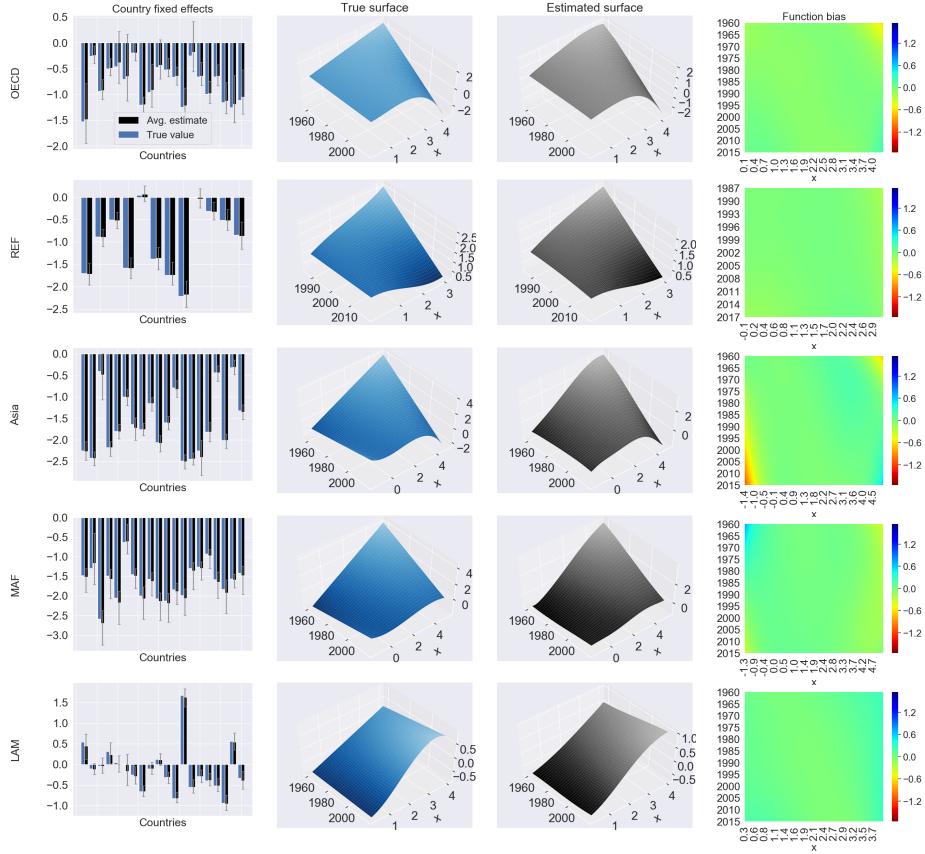


Note: In the country fixed effects plots, gray error bars indicate 95% confidence intervals.

for estimation. The boundary issues seem most severe for the linear-cubic function, likely because most of the cubic curvature becomes the most pronounced toward the boundaries of the input region.

Figures 1.9 | 1.11 summarize average estimates across Monte Carlo samples for the regional formulation of the dynamic model for the linear, linear-quadratic, and linear-cubic function, respectively. Despite large variations in the number of observations available for each region, the model performs well for each region and for every function considered.

**Figure 1.11:** Monte Carlo results for the regional formulation of the dynamic model: linear-cubic function

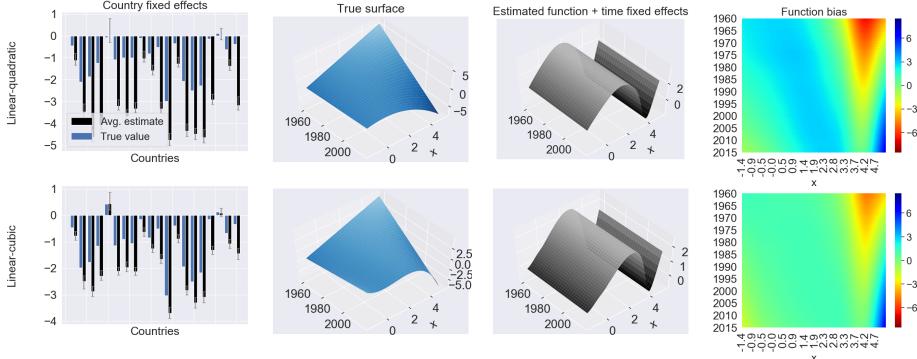


Note: In the country fixed effects plots, gray error bars indicate 95% confidence intervals.

### 1.5.3 Misspecification

In this section, we investigate effects of model misspecification on our estimator of the parameters in the static model. We are interested in performance of the estimator of the static model in cases where the true income-emissions relationship is time varying. We simulate observations using the “linear-quadratic” and “linear-cubic” functions presented above, then estimate a global and regional ( $R=5$ ) formulation of the static model. The underlying simulations are thus the same as those used in Section 1.5.2 but we now apply the estimation method used in Section 1.5.1. For ease of presentation, we focus on results for the global model formulation. We report optimal network architectures based on an initial Monte Carlo sample in Table A.6 of the appendix.

Note while the static model estimates a one-dimensional income-emissions

**Figure 1.12:** Monte Carlo misspecification results for the global formulation of the static model

Note: In the country fixed effects plots, gray error bars indicate 95% confidence intervals.

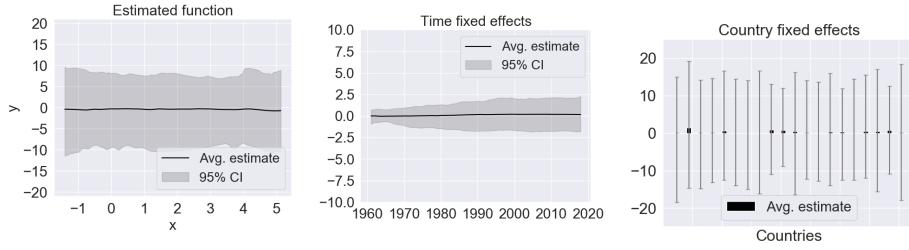
relationship, we can construct an implied, two-dimensional, estimated income-emissions surface by adding estimated time fixed effects to the estimated income-emissions function. Figure 1.12 shows estimation results for the global model formulation. Results for the regional model formulation are similar and are reported in Figures A.1 and A.2 of the appendix. The first column in Figure 1.12 shows true and estimated country fixed effects; the second column shows the true income-emissions surfaces; the third column shows the estimated income-emissions surfaces (estimated functions plus estimated time fixed effects); and the fourth column shows function bias obtained as the difference between the surfaces in the second and third column. The results in Figure 1.12 can be compared to Figure 1.8 which contains estimation results from the same underlying data, but where we estimate the income-emissions relationship using the dynamic model.

Comparing Figures 1.8 and 1.12 it is clear the static model is not able to capture the time-varying income-emissions relationships. The implied, estimated surfaces from the static model (Figure 1.12) have a much higher bias than those of the dynamic model (Figure 1.8). This shows that if the income-emissions relationship is sufficiently varying in time, the static model with time fixed effects is not able to capture this. In this case, it is important to use a more flexible model such as the dynamic model proposed above.

#### 1.5.4 Spuriousness

In this section, we investigate whether our estimator of the parameters in the static and the dynamic model suffers from spuriousness in settings with independent but stochastically trending variables. Specifically, we simulate independent random walks to be used as the dependent variable in the models:

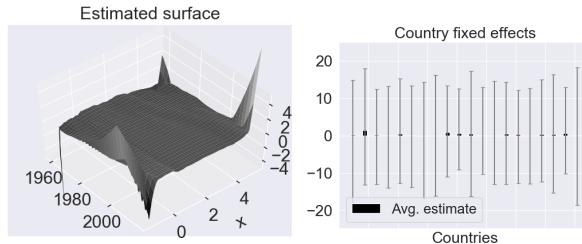
$$y_{it} = y_{it-1} + \nu_{it}, \quad i = 1, \dots, N_t, \quad t = 1, \dots, T, \quad (1.12)$$

**Figure 1.13:** Monte Carlo spuriousness results for the global formulation of the static model

Note: In the country fixed effects plot, gray error bars indicate 95% confidence intervals.

where  $y_{i0} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$  and  $v_{it} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_v^2)$ . We then estimate a global and a regional ( $R=5$ ) formulation of both the static and the dynamic model using the methodology proposed above. That is, we estimate models where the dependent variable  $y_{it}$  is simulated from (1.12), and the input variable  $x_{it}$  is the natural logarithm of per capita GDP from the CO<sub>2</sub> emissions panel discussed in Section 1.3. Clearly,  $y_{it}$  and  $x_{it}$  are independent and characterized by country-specific stochastic trends. For ease of presentation, we again focus on results for the global model formulation. We report optimal network architectures based on an initial Monte Carlo sample in Table A.7 of the appendix.

Figures 1.13 and 1.14 show estimation results for the global formulation of the static and the dynamic model, respectively. For both models, the average estimate of the income-emissions relationship is around zero throughout the input region used for estimation, as it should be since the dependent variable and the input variable are completely unrelated. Similarly, average estimates of all fixed effects are around zero, as they should be. Results for the regional model formulation are similar and are reported in Figures A.3 and A.4 of the appendix. The results of this section suggest our estimator does not suffer from spuriousness and can be reliably used in practice. The results are perhaps not that surprising, as it is well-known that spuriousness is

**Figure 1.14:** Monte Carlo spuriousness results for the global formulation of the dynamic model

Note: In the country fixed effects plot, gray error bars indicate 95% confidence intervals.

less problematic in panel data whenever the time dimension grows faster than the cross-sectional dimension [Levin, Lin, and Chu 2002].

## 1.6 Empirical Analysis

This section presents the results of our empirical analysis on the data introduced in Section 1.3. In Section 1.6.1 we use traditional territorial emissions, while Section 1.6.2 employs consumption-based emissions as the dependent variable.

### 1.6.1 Territorial emissions

To investigate the income-emissions relationship, we estimate a global and a regional formulation of both the static and the dynamic neural network model, as well as a national formulation of the static model. Optimal network architectures for each model, obtained by minimizing the BIC in 1.11, are presented in Table 1.4. As previously discussed, the learning task faced by the regional model formulation is more complex than that faced by its global counterpart, and similarly, the learning task faced by the national model formulation can be considered more complex than that faced by its regional counterpart. Likewise, the learning task faced by the dynamic model is in many ways also more complex than that faced by its static counterpart. From Table 1.4 we see a clear tendency that increased complexity of the learning task prompts increased complexity of the neural network component. In particular, depth in the network architecture appears to be important for the dynamic model specification. The network architecture used for the regional formulation of the dynamic model is the most complex. The national formulation of the static model may be regarded as the least parsimonious of the models considered, as it contains the largest number of free model parameters excluding fixed effects. However, simply comparing the number of free model parameters excluding fixed effects across the static and the dynamic model specification, and across the three model formulations (global, regional and national) is misleading. For instance, the

**Table 1.4:** Optimal neural network architectures for territorial emissions

	Static model			Dynamic model	
	Global	Regional	National	Global	Regional
Network architecture	(2, 2)	(8)	(2, 2, 2)	(4, 4, 2)	(8, 8, 8)
# parameters (excl. fixed effects)	12	56	388	44	208

Note: “(a,b,c)” indicates a neural network architecture with three hidden layers containing *a* units in the first layer, *b* in the second, and *c* in the third.

**Table 1.5:** Fractions of variance explained for territorial emissions

	Global	OECD	REF	Asia	MAF	LAM
Country effects only	0.92	0.79	0.86	0.81	0.91	0.85
Time effects only						
Global	0.04	-6.63	-0.52	-0.03	-0.23	0.05
Regional	0.38	0.04	0.07	0.11	0.05	0.10
Full static model						
Global	0.96	0.87	0.86	0.95	0.95	0.93
Regional	0.97	0.90	0.96	0.96	0.96	0.94
National	0.98	0.96	0.92	0.98	0.97	0.96
Full dynamic model						
Global	0.97	0.89	0.86	0.96	0.96	0.93
Regional	0.98	0.94	0.97	0.97	0.97	0.95

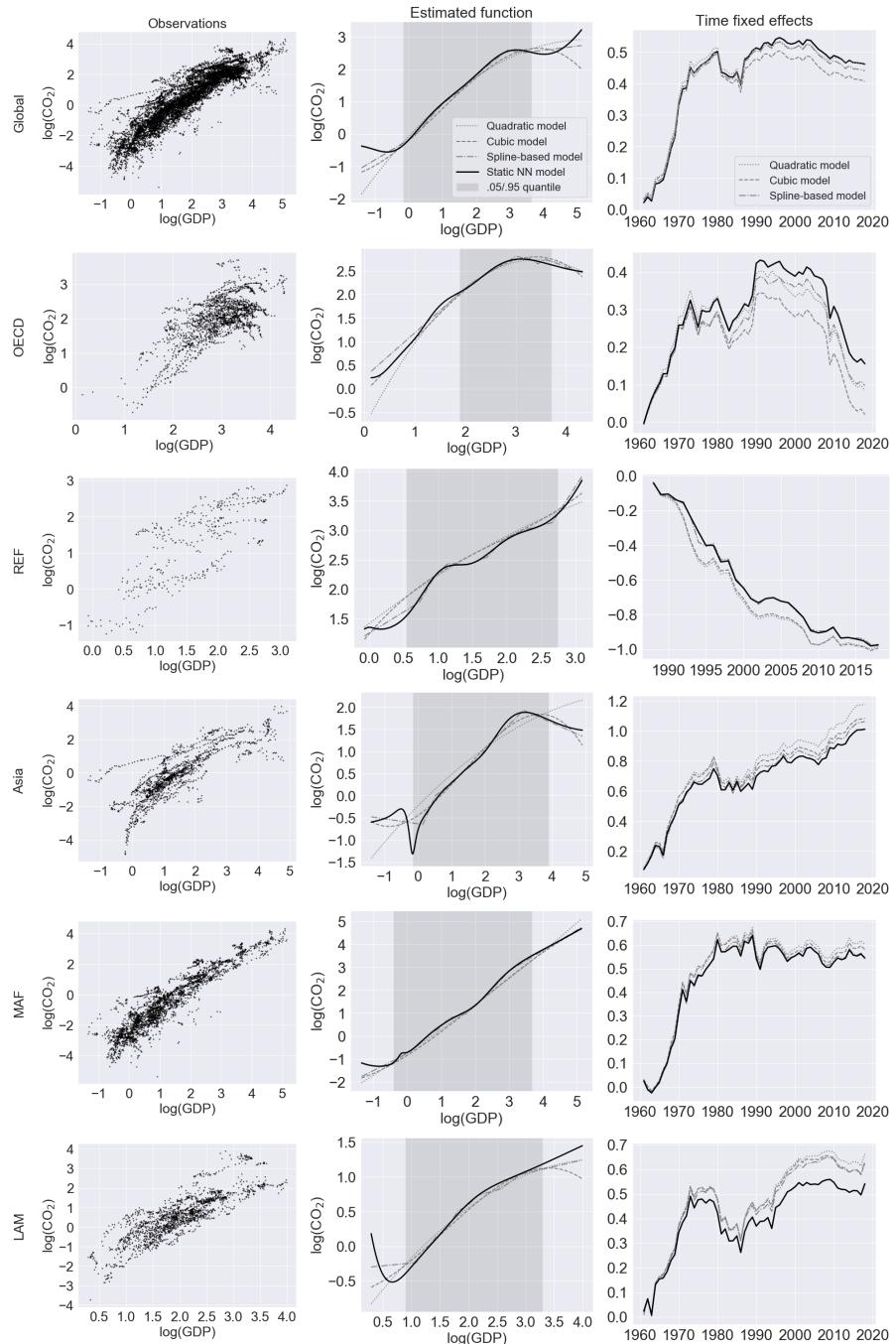
Note: Numbers in the table are  $R^2$  statistics calculated using observations belonging to a given region. Global  $R^2$  statistics are calculated using all observations. Line one is obtained by a regression using only country dummies. Line two is obtained by a regression using only time dummies. Line three is obtained by a regression in which time dummies have been interacted with a regional indicator. The last lines follow from estimation of the full models including income effects, country fixed effects, and potentially time fixed effects. We have highlighted the globally preferred formulation of each model.

static model specification requires estimation of time fixed effects, while the dynamic model specification does not. Also, the national formulation of the static model contains many country-specific slope coefficients in the output layer, but does not require estimation of any fixed effects. Taking into account fixed effects, the regional formulation of the static model incurs most parameters, a total of 500. Likewise, when taking into account fixed effects, the global and the regional formulations of the static model contain more free parameters than their dynamic counterparts.

Table 1.5 provides information on how much sample variance is explained by the different model components, see the caption of the table for details. The first column of Table 1.5 shows how much of the global sample variance is explained, remaining columns show how much of the region-specific sample variance is explained by the respective models. We find that country fixed effects alone explain as much as 92% of the global sample variance, and that global time fixed effects explain only 4% of the global sample variance. So even though our data set spans six decades, cross-country differences are more important than within-country differences. Similar conclusions were reached in Schmalensee et al. (1998). Note also the fraction of variance explained varies across regions. In particular, we note a negative  $R^2$  in the

second line of Table 1.5 for the regions OECD, REF, Asia, and MAF, suggesting global time fixed effects fit these regions worse than a region-specific sample average. From the third line of Table 1.5 we see that allowing for region-specific time fixed effects helps increase the fraction of variance explained. This is true for all regions as well as globally. This is of course not surprising given that we found cross-country differences to be important. The last five lines of Table 1.5 show we can improve the fraction of variance explained by country fixed effects by between 6 and 17 percentage points, depending on the region, by specifying one of our proposed models. For both the static and the dynamic model specification, we see that a regional model formulation improves the fraction of variance explained by a global model formulation. This is true globally as well as for each region separately. Similarly, for the static model specification, we note that a national model formulation generally improves upon the fraction of variance explained by the regional model formulation, globally, as well as for each region separately. The only exception is the region REF. We also note that for a given formulation (global or regional), the dynamic model never performs worse than the static model. Comparing the best performing formulations of the static and the dynamic model specifications, the national formulation of the static model and the regional formulation of the dynamic model appear to perform on par. Indeed, they explain the same amount of global sample variance to two decimals, and they explain the same amount of region-specific sample variance for the region MAF. The national formulation of the static model explains most region-specific sample variance for three out of the remaining four regions, and the regional formulation of the dynamic model explains most sample variance for the fourth region.

We show estimation results for the global and the regional formulation of the static neural network model in Figure 1.15. From the discussion in Section 1.5 recall that we are most confident about the model output within the interior of the input region used for estimation not too close to the boundaries. In Figure 1.15 we therefore shade the area between the .05 and the .95 income quantiles when we plot the estimated income-emissions relationship to indicate the area of highest confidence. The first row of plots in Figure 1.15 shows estimation results for the global model formulation. For reference, we also include estimation results from popular models within the EKC literature: a quadratic EKC model (Holtz-Eakin and Selden 1995), a cubic EKC model (Grossman and Krueger 1991), and a spline-based EKC model (Schmalensee et al. 1998). The benchmark models include country and time fixed effects like the static neural network model but employ different specifications of the emissions function. Remaining rows in Figure 1.15 show estimation results for the regional formulation of the static neural network model. For reference, we include results from region-wise estimation of the benchmark models. Note that region-wise estimation of the benchmark models ignores cross-region dependencies. By contrast, the regional formulation of the neural network models proposed in this paper uses information across regions to learn common input transformations.

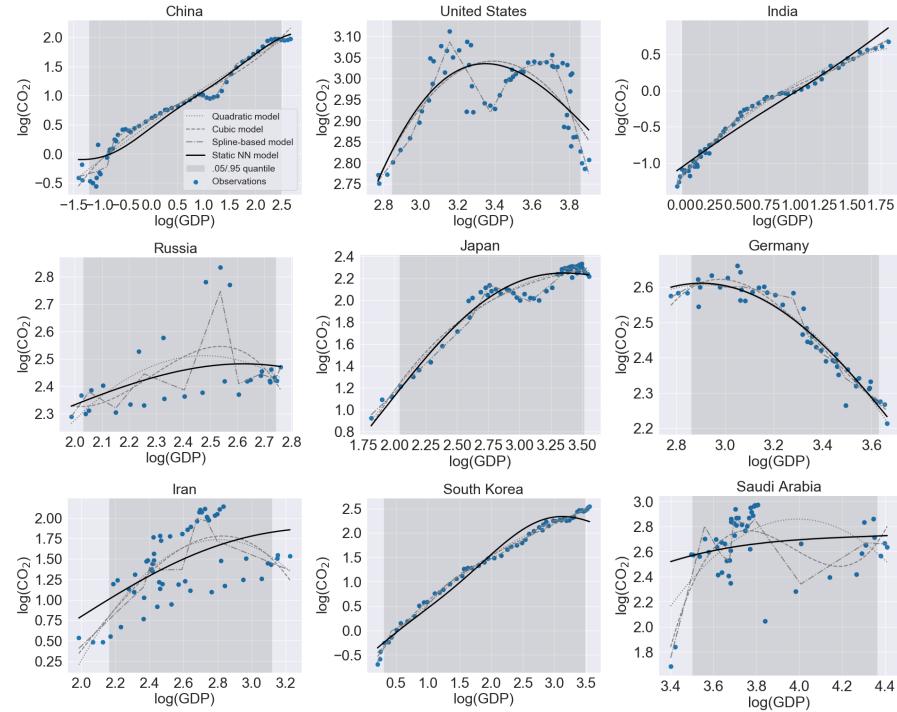
**Figure 1.15:** Estimation results for the static model and territorial emissions

Note: Estimation results in the first row of plots are based on the global formulation of the static neural network model. Remaining rows are based on the regional formulation of the static neural network model. Benchmark models from the EKC literature are included in gray for reference.

Inspecting Figure 1.15 we initially see evidence of a global EKC relationship for territorial CO<sub>2</sub> emissions. For the global model formulation in the first row of the figure, we see evidence of an EKC-shape in both the estimated income-emissions relationship and in the estimated time fixed effects. From the regional model formulation, we also note evidence of an EKC-shape for OECD and Asia. For OECD, the static model shows a clear EKC-shape in both the estimated income-emissions relationship as well as in the estimated time fixed effects. For Asia, the static model again shows an EKC-shape in the estimated income-emissions relationship. For Asia, however, the EKC-shape is less clear in the estimated time fixed effects than it is for OECD. It is interesting that we observe an EKC-shape for both OECD and Asia. Recall from the discussion in Section 1.3 that OECD is the richest of the five regions considered, while Asia is one of the poorest. Therefore, one could have imagined that only countries within OECD would have become rich enough to reach a turning point. However, this does not seem to be the case. For the remaining regions REF, MAF, and LAM, the static model shows a monotonically increasing income-emissions relationship. The estimated time fixed effects display some more concavity for these regions, and even a downward trend for the region REF. However, there is no evidence of an EKC-shape for these regions. Notice the sample period does not start until 1990 for the region REF, and so the downward trending time fixed effects observed for this region in fact correspond to the period for which we also observe a downward trend in the time fixed effects for the regions OECD and MAF as well as for the global time fixed effects. Generally, results from the benchmark models appear in line with our static neural network model.

We use the national formulation of the static model to estimate country-specific income-emissions relationships. Figure 1.16 shows estimates of the country-specific, income-emissions relationships for the world's nine largest CO<sub>2</sub> emitting countries. For reference, we also include results from country-wise estimation of the benchmark models. By using cross-country information to learn common input transformations, it seems our static neural network model is generally able to learn more stable income-emissions relationships that are easier to interpret than the benchmark models. In Table 1.6 we report the number of estimated, country-specific EKCs within each region using our static neural network model, defined as an inverse U-shaped relationship with an in-sample turning point, as well as the number of strictly increasing, strictly decreasing, and other shapes of the income-emissions relationship. From Table 1.6 we see the most common shape across regions is the strictly increasing and the second most common shape is the EKC. We observe the most EKCs among OECD, which also has the largest ratio of EKCs to non-EKC-shapes. For Asia, the ratio of EKCs to non-EKC-shapes is considerably lower than for OECD. Yet, we observe a region-specific EKC for both of these regions when estimating a regional model formulation. Thus, it seems the regional EKC observed for Asia might be driven only by few wealthy countries such as Japan and South Korea, cf. Figure 1.16 Interest-

**Figure 1.16:** Estimated income-emissions relationship for the world's nine largest CO<sub>2</sub> emitters using territorial emissions data



Note: Estimated income-emissions relationships are based on the national formulation of the static neural network model. Benchmark models from the EKC literature are included in gray for reference.

ingly, the ratio of EKCs to non-EKC-shapes is also quite large for REF. However, this does not result in a region-specific EKC for REF when estimating a regional model formulation. This may be because the country-specific EKCs among REF are much less distinct than those among the OECD and Asia. The EKC for Russia in Figure 1.16 is representative of the EKCs among REF.

The dynamic neural network model allows us to take the analysis one step further than what is traditionally done in the literature. It allows us to investigate how the income-emissions relationships potentially change shape over time. We show estimation results for the dynamic neural network model in Figure 1.17. In the top left panel of the figure, we show the estimated surface from the global model formulation. Remaining panels show estimated surfaces from the regional model formulation. We indicate the area between the .05 and the .95 income quantiles by coloring the estimated surface black. Note this area is now varying with time. We also indicate the area outside of this time-varying quantile range but within the minimum and

**Table 1.6:** Country-specific shapes of the income-emissions relationship within each region for territorial emissions

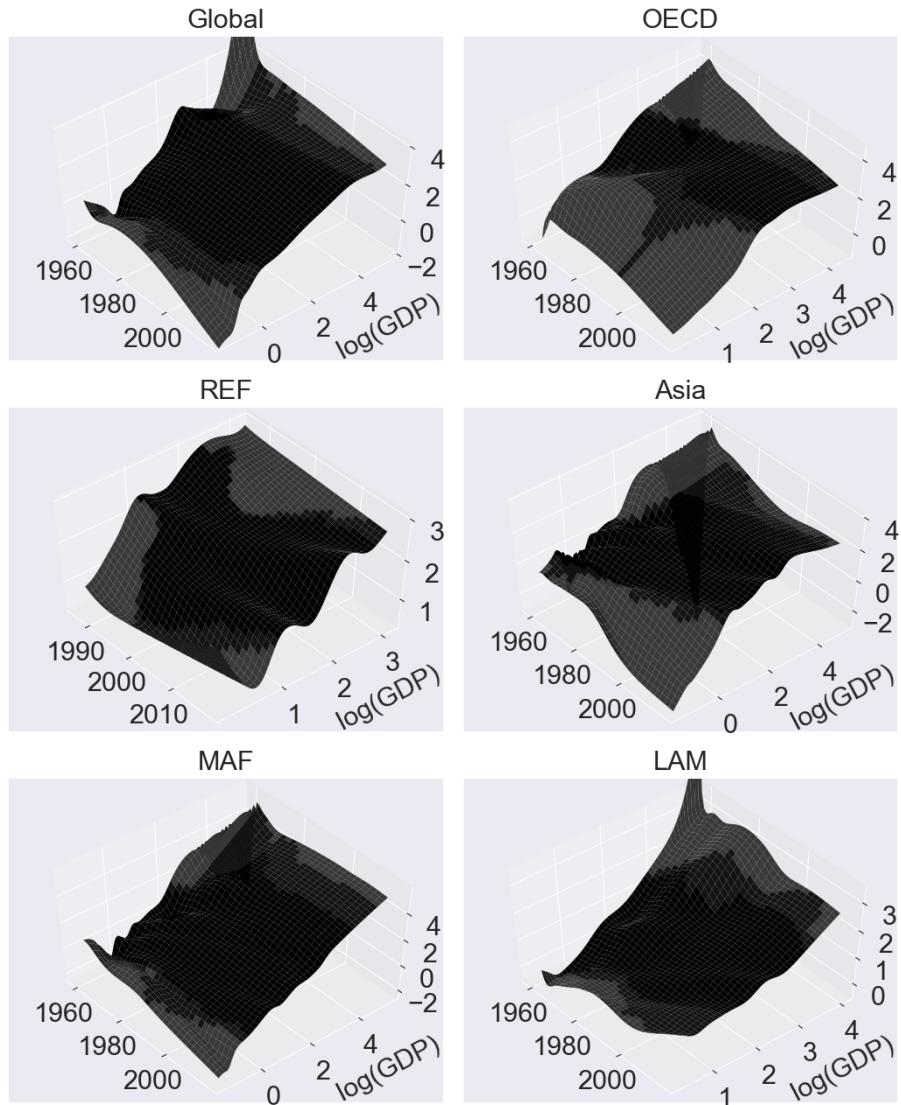
	OECD	REF	Asia	MAF	LAM	Total
EKC	25	4	5	4	3	41
Increasing	10	6	29	53	28	126
Decreasing	6	3	1	5	3	18
Other	0	0	0	1	0	1

Note: The first row of the table reports the number of estimated EKC relationships within each region, defined as an inverse U-shaped relationship with an in-sample turning point, based on the national formulation of the static neural network model. The second and third rows of the table report the number of strictly increasing and decreasing estimated income-emissions relationship within each region, respectively. The fourth row of the table (“Other”) represents Mozambique, which has an estimated U-shaped income-emissions relationship.

maximum income value observed for a given time period by coloring the estimated surface dark grey. The estimated surface is colored light grey over areas of the input space where the model has no data support.

Similar to the static model, the global formulation of the dynamic model shows evidence of a global EKC in the top left panel of Figure 1.17. We also note the global EKC-shape appears rather stable throughout the sample period. By contrast, it appears from the estimated surfaces using the regional formulation of the dynamic model that the EKC-shapes observed for OECD and Asia do not appear in the data until late in the sample period. Like the static model, also the dynamic model suggests a monotonically increasing income-emissions relationship for the regions REF, MAF, and LAM. The surface plot for the region REF is somewhat noisy, but, overall, there seems to be a monotonically increasing income-emissions relationship. The monotonically increasing relationships observed for the regions MAF and LAM appear rather stable over time.

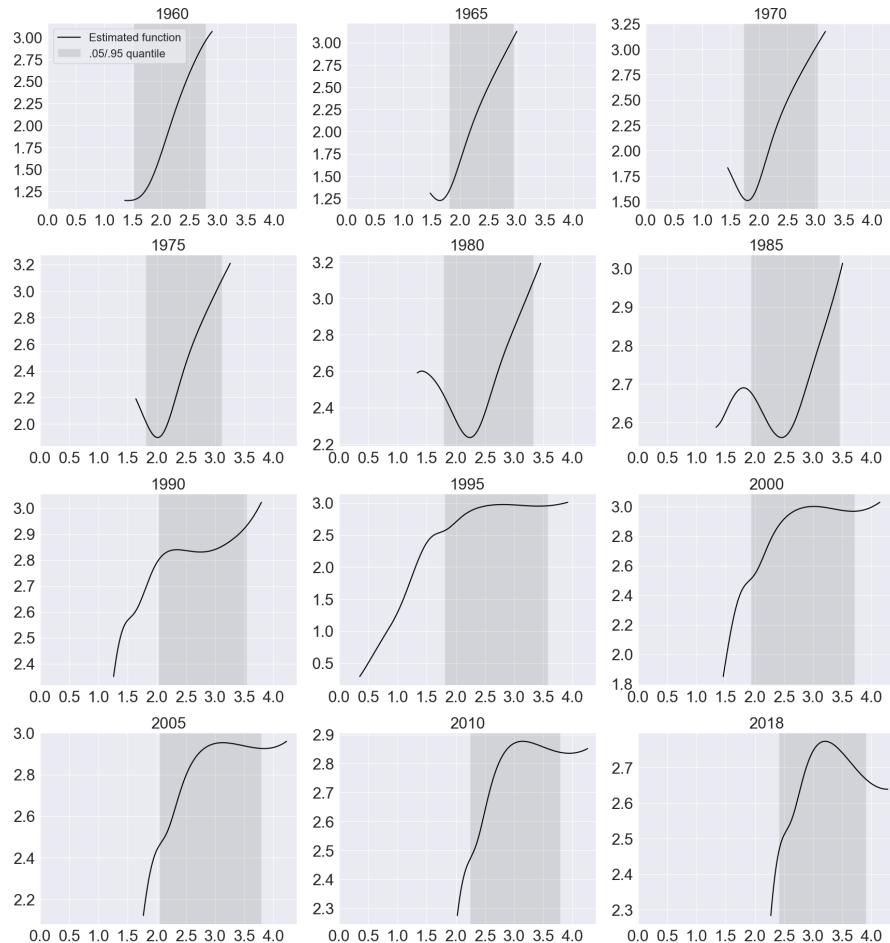
In Figure 1.18, we illustrate clearly how the income-emissions relationship for OECD changes shape over time by “slicing” the estimated surface for OECD every five years. We note from Figure 1.18 that the EKC-shape does not show up in the data for OECD until the early to mid 1990s. The turning point for OECD occurs around a value of log per capita GDP equal to 3, which is similar to what we find globally and for Asia in Figure 1.15 and is supported by the benchmark models. However, it is interesting to note from Figure 1.18 that OECD reaches levels of log per capita GDP above 3 already in the 1960s, and log per capita GDP values above 3 is within the area between the .05 and .95 quantiles since the 1970s. Therefore, it seems a structural shift might occur around the early to mid 1990s that allows countries with high incomes to emit less CO<sub>2</sub>. This shift could be the result of increased environmental concerns,

**Figure 1.17:** Estimation results for the dynamic model and territorial emissions

Note: Black coloring indicates area between the .05 and .95 income quantiles. Dark gray coloring indicates area with data support. Light grey coloring indicates area with no data support. The estimated surface in the top left panel is based on the global formulation of the dynamic neural network model. Remaining surfaces are based on the regional formulation of the dynamic neural network model.

climate change mitigation policies, or other factors unrelated to income. For instance, the Montreal Protocol of the United Nations was adopted in 1987 with the goal of

**Figure 1.18:** Time-varying environmental Kuznets curve for OECD using territorial emissions data



Note: Estimated income emissions relationships are based on the regional formulation of the dynamic neural network model.

protecting the stratospheric ozone layer by phasing out production and consumption of ozone-depleting substances, and the United Nations Framework Convention on Climate Change (UNFCCC) was established in 1994. The shift could also reflect the pollution haven hypothesis that we discuss below.

### 1.6.2 Consumption-based emissions

We now investigate to what extend the income-emissions relationships, documented in the previous section, are driven, or affected, by international trade patterns. We

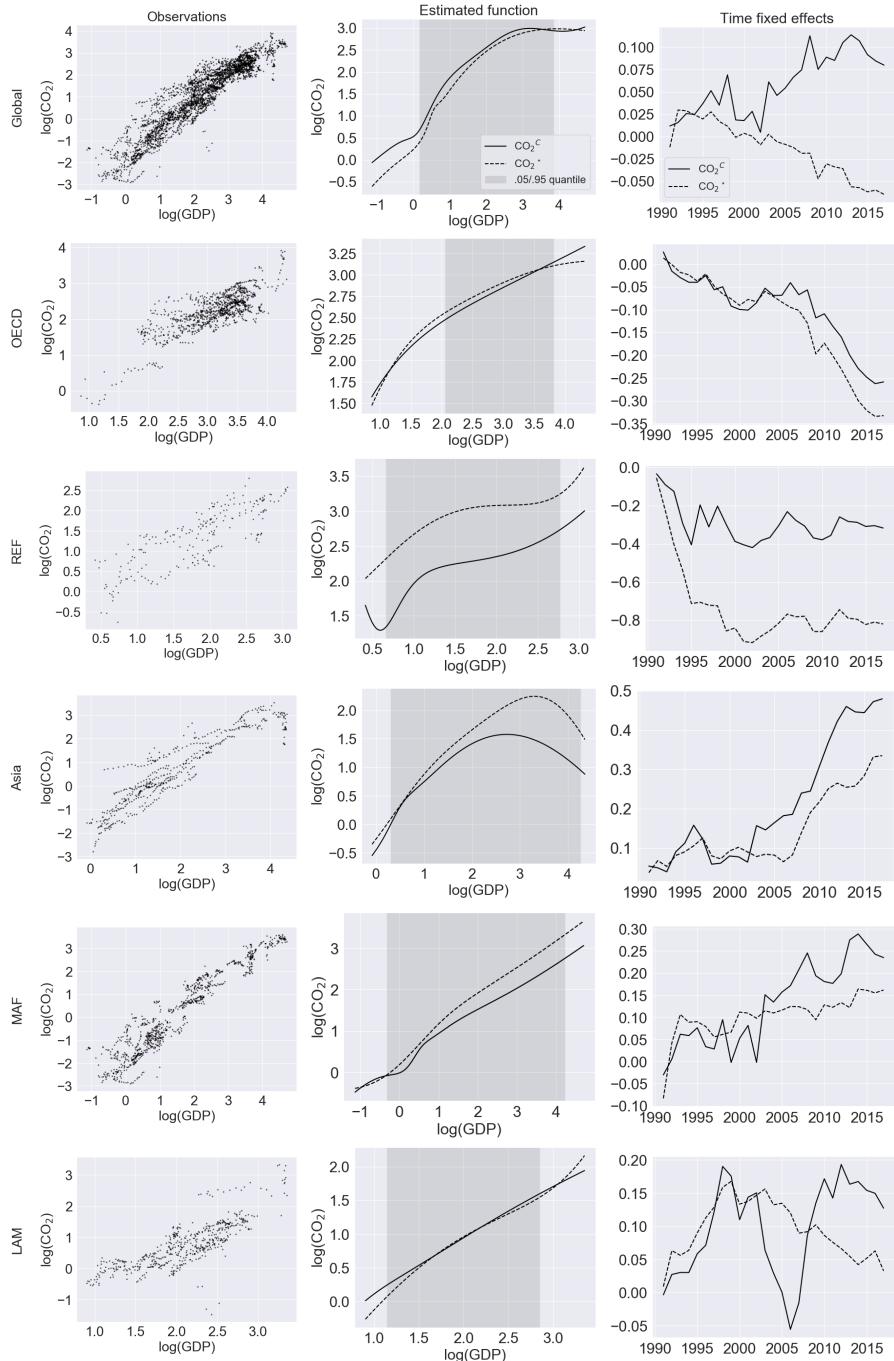
**Table 1.7:** Optimal neural network architectures for consumption-based emissions

	Static model		Dynamic model	
	Global	Regional	Global	Regional
$\text{CO}_2^C$				
Neural network architecture	(2, 2)	(4)	(4, 2)	(8, 4)
# parameters (excl. fixed effects)	12	28	24	80
$\text{CO}_2^*$				
Neural network architecture	(2, 2)	(4)	(4, 2, 2)	(8, 8, 8)
# parameters (excl. fixed effects)	12	28	30	208

Note: " $\text{CO}_2^C$ " are consumption-based CO<sub>2</sub> emissions; " $\text{CO}_2^*$ " are territorial CO<sub>2</sub> emissions based on a restricted sample that coincides with the one for consumption-based emissions; "( $a, b, c$ )" indicates a neural network architecture with three hidden layers containing  $a$  units in the first layer,  $b$  in the second, and  $c$  in the third.

do so by reestimating the global and the regional formulation of the static and the dynamic neural network model using consumption-based emissions for the dependent variable as opposed to territorial emissions; see Section 1.3 for the definition of consumption-based emissions. Since consumption-based emissions are available only from year 1990 onward, leaving the time series available for each country rather short, we do not estimate a national formulation of the static model. To facilitate comparison between the results obtained from using territorial- versus consumption-based emissions, we also reestimate all models for territorial emissions restricting the sample to coincide with the sample for consumption-based emissions. This shows whether observed changes in the income-emissions relationships can plausibly be attributed to the switch from territorial emission to consumption-based emissions, or whether changes are simply an effect of a restricted data set.

In the first line of Table 1.7, we report optimal neural network architectures for consumption-based emissions, obtained by minimizing the BIC (1.11). In the third line of Table 1.7 we report optimal neural network architectures for territorial emissions using the restricted sample. Overall, Table 1.7 is broadly comparable to Table 1.4 of the previous section. Namely, a more complex network architecture is required for the dynamic model than for the static model, and a more complex network architecture is required for a regional model formulation than for a global model formulation. We also note that, for the dynamic model, a higher degree of model complexity is required for territorial emissions on the restricted sample than for consumption-based emissions. This is especially true for the regional model formulation. This provides an indication that perhaps the region-specific income-emission relationships for consumption-based emissions are less complex than those for territorial emissions.

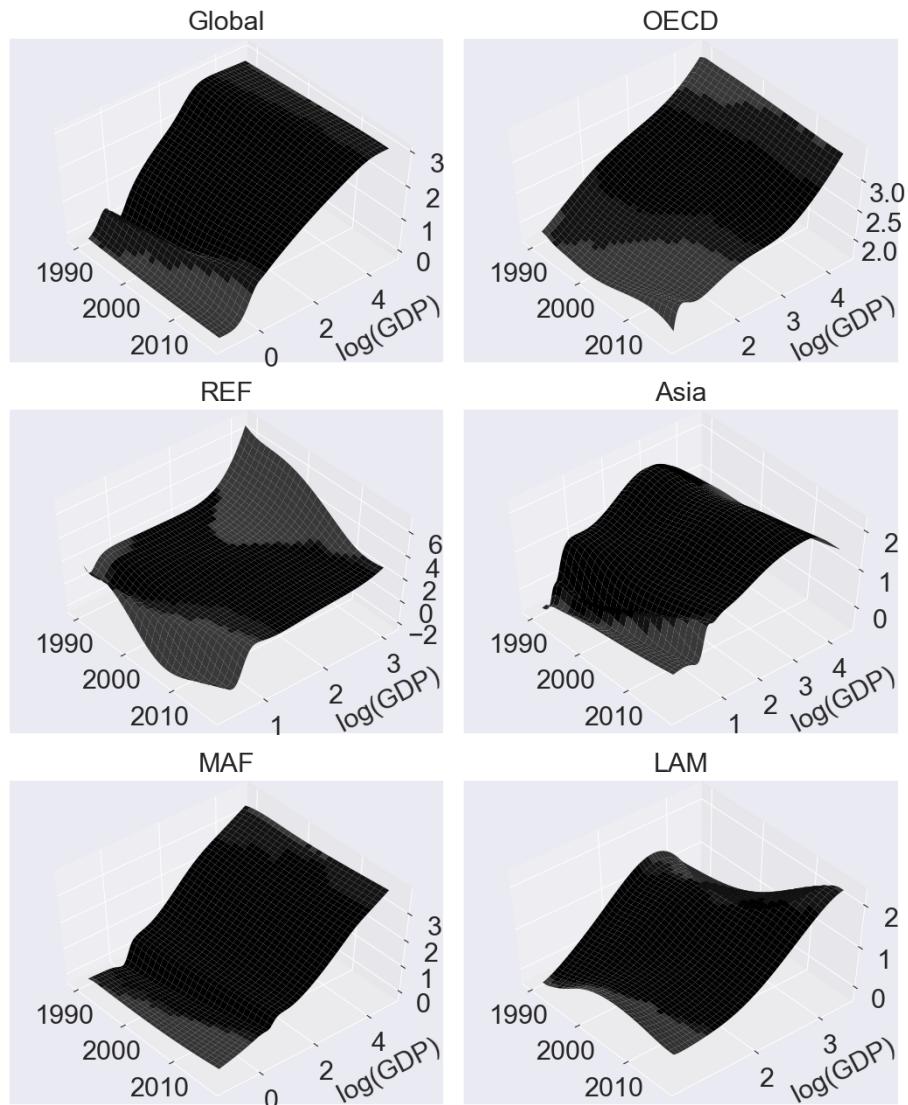
**Figure 1.19:** Estimation results for the static model and consumption-based emissions

Note: "CO<sub>2</sub><sup>C</sup>" are consumption-based CO<sub>2</sub> emissions and "CO<sub>2</sub><sup>\*</sup>" are territorial CO<sub>2</sub> emissions based on a restricted sample that coincides with the one for consumption-based emissions. Estimation results in the first row of plots are based on the global formulation of the static neural network model. Remaining rows are based on the regional formulation of the static neural network model.

Figure 1.19 shows estimation results for the static neural network model for consumption-based emissions as well as territorial emissions based on the restricted sample. The first row of plots in the figure shows estimation results for the global formulation of the static model. Remaining rows show estimation results for the regional formulation of the static model. In plots of the estimated income-emissions relationship (second column) and time fixed effects (third column), a solid line indicates results for consumption-based emissions, and a dashed line indicates results for territorial emissions based on the restricted sample. Estimation results for the quadratic, the cubic, and the spline-based benchmark models using consumption-based emissions for the dependent variable can be found in Figure A.5 of the appendix. They are generally in line with the results from the static neural network model in Figure 1.19.

Inspecting the results from the global formulation of the static model in the first row of Figure 1.19, we note a downward trend in estimated time fixed effects for territorial emissions. However, recall the sample does not start until year 1990. This implies the downward trend in estimated time fixed effects for territorial emissions in the top right panel of Figure 1.19 corresponds to the downward trend in estimated time fixed effects for that same period in the top right panel of Figure 1.15. Conversely, estimated time fixed effects for consumption-based emissions are trending upwards in the top right panel of Figure 1.19. But, we also note that the magnitude of the estimated time fixed effects over this period is small, suggesting the effect of time on the global income-emission relationship is rather minimal over this sample period. In the top middle panel of Figure 1.19, we note the estimated income-emissions relationship from the global model formulation exhibits concavity for both types of emissions in line with the EKC hypothesis. That said, the global EKC-shape is not quite as clear as it is in Figure 1.15. We conclude that when controlling for international trade flows, we still see concavity in the global income-emissions relationship, i.e. the EKC relationship appears intact.

Inspecting the results from the regional formulation of the static model in the remaining rows of Figure 1.19, we immediately note the previously observed EKC-shape for OECD when using territorial emissions disappears when using consumption-based emissions. The income-emissions relationship appears to be strictly increasing for OECD when using consumption-based emissions and thus not reaching a turning point or even flattening out. Although the EKC-shape is not quite as clear for territorial emissions based on the restricted sample as it was for the full sample in Figure 1.15, there still is pronounced concavity for territorial emissions based on the restricted sample. Recall from the discussion in Section 1.3 that OECD is the only region that is a net-exporter of CO<sub>2</sub> emissions to other countries throughout the restricted sample period from year 1990 onward. Together with the fact that the EKC-shape disappears for OECD when using consumption-based emissions this suggests that the apparent EKC relationship we found when using territorial emissions plausibly can be explained by the pollution haven hypothesis. Recall the pollution

**Figure 1.20:** Estimation results for the dynamic model and consumption-based emissions

Black coloring indicates area between the .05 and .95 income quantiles. Dark gray coloring indicates area with data support. Light grey coloring indicates area with no data support. The estimated surface in the top left panel is based on the global formulation of the dynamic neural network model. Remaining surfaces are based on the regional formulation of the dynamic neural network model.

haven hypothesis suggests that rich countries, in this case belonging to the OECD region, simply export their emissions to other countries, often poor countries subject

to weak environmental regulation, through international trade. For Asia, we note, by contrast, that the EKC-shape becomes even crisper and with an earlier turning point when using consumption-based emissions than when using territorial emissions. We interpret this as the mirror image of the result for OECD and as further support for the pollution haven hypothesis; recall also from the discussion in Section 1.3 that Asia is a net importer of CO<sub>2</sub> emissions throughout the restricted sample period. For most of the remaining regions, the shape of the estimated income-emissions relationship does not change substantially when moving from territorial emissions to consumption-based emissions. For REF, there is a substantial level shift in the income-emissions relationship estimated by the static model when moving from territorial emissions to consumption-based emissions. However, we see a similar level shift in the estimated time fixed effects, and so this level shift is most likely just the result of an identification issue stemming from the small number of observations available for this region.

We show estimation results from the global and the regional formulation of the dynamic neural network model when using consumption-based emission for the dependent variable in Figure 1.20. Estimation results from the dynamic model when using territorial emissions based on the restricted sample for the dependent variable can be found in Figure A.6 of the appendix; they look similar to the results from the dynamic model when using territorial emissions based on the unrestricted sample for the dependent variable in Figure 1.17. Inspecting the estimated surface in the top left panel of Figure 1.20 obtained from the global model formulation, it shows concavity in the income dimension, and it seems stable in the time dimension. As such, the conclusion from the global formulation of the static model of a global EKC also for consumption-based emissions seems stable over the restricted sample. Inspecting the estimated surfaces from the regional model formulation in the remaining panels of Figure 1.20, we note also the conclusions from the regional formulation of the static model appear stable over the restricted sample. In particular, the strictly increasing income-emissions relationship observed for OECD when using consumption-based emissions for the dependent variable appear stable over the restricted sample, and the EKC-shape observed for Asia when using consumption-based emissions for the dependent variable appear stable over the restricted sample.

## 1.7 Conclusion

In this paper, we propose a novel neural network-based panel data methodology for analyzing the environmental Kuznets (EKC) curve for carbon dioxide (CO<sub>2</sub>) emissions. We consider two distinct model specifications within this overall framework: a static model specification and a dynamic model specification. The static model consists of both country and time fixed effects in addition to a feedforward neural network component with income as the only input variable. This model is static in

the sense that the shape of the income-emissions relationship is assumed to be fixed over time. The dynamic model uses a time variable as an additional input into the neural network component, in place of time fixed effects. By using both income and a time variable as inputs into the neural network component, the dynamic model is able to learn how time and income interact, and how the income-emissions relationship potentially changes its entire shape over time. Both model specifications use cross-country dependencies to learn common input transformations by having some model parameters be shared across all countries, and simultaneously allow for cross-country heterogeneity in the shape of the income-emissions relationship by having other parameters be specific to regions of homogeneous countries. In a Monte Carlo study, we demonstrate that our proposed methodology is able to identify various functional forms of different complexity, and we also demonstrate its ability to account for country-specific stochastic trends. In addition, we demonstrate that when the simulated income-emissions relationship is sufficiently time varying, the static model with time fixed effects is not able to capture this, but the dynamic model is able to appropriately capture time-varying income-emissions relationships.

We investigate the relationship between per capita GDP and CO<sub>2</sub> emissions, using national-level panel data for the period 1960-2018. When using territorial emissions data, we find that, globally, as well as for the regions OECD and Asia, there is evidence of an inverse U-shaped income-emissions relationship, often referred to as an environmental Kuznets curve. Globally, this relationship appears rather stable over time. However, for both OECD and Asia, it seems the EKC-shape does not appear until late in the sample period. Conversely, when using consumption-based emissions, a measure of emissions that accounts for cross-country emissions transfers through international trade, the picture changes substantially. The evidence of an EKC relationship for OECD dilutes, while, for Asia, the EKC-relationship becomes clearer and with an earlier turning point. This suggests the apparent EKC-relationship observed for OECD when using territorial emissions is driven by emissions exports to other countries. Conversely, it also suggests that, for Asia, there are EKC-effects in emissions due to local consumption that are not seen in territorial emissions due to imported emissions.

There are many directions in which our work could be extended in the future. For instance, we believe the methodology proposed in this paper can be useful in a wide range of different settings, such as for prediction of panels of economic data. In a follow-up paper [Bennedsen, Hillebrand, and Jensen 2021], we extend the methodology proposed in this paper by considering a long short-term memory (LSTM) recurrent neural network architecture, and construct scenario-based CO<sub>2</sub> emissions projections through 2100 using scenarios from the Shared Socioeconomic Pathways [Riahi et al. 2017].

## Acknowledgements

For useful comments and suggestions, we thank Professor Siem Jan Koopman and participants at EGU2020: Sharing Geoscience online (2020), the Joint Econometrics-Finance seminar (2020) at the Center for Research in Econometric Analysis of Time Series (CREATES) at Aarhus University, the EDS Brownbag semiar (2021) at Vrije Universiteit Amsterdam, and the Economics and Business Economics Seminar (EBA) at Aarhus University (2021). The authors also acknowledge financial support from the DFF grant Econometric Modeling of Climate Change.

## 1.8 References

- Agras, J., Chapman, D., 1999. A dynamic approach to the Environmental Kuznets Curve hypothesis. *Ecological Economics* 28 (2), 267 – 277.  
URL <http://www.sciencedirect.com/science/article/pii/S0921800998000408>
- Anders, U., Korn, O., 1999. Model selection in neural networks. *Neural Networks* 12 (2), 309 – 323.  
URL <http://www.sciencedirect.com/science/article/pii/S0893608098001178>
- Arrow et al., 1995. Economic growth, carrying capacity, and the environment. *Ecological Economics* 15 (2), 91 – 95.  
URL <http://www.sciencedirect.com/science/article/pii/0921800995000593>
- Aslanidis, N., Sep. 2009. Environmental Kuznets Curves for Carbon Emissions: A Critical Survey. Working Papers 2009.75, Fondazione Eni Enrico Mattei.  
URL <https://ideas.repec.org/p/fem/femwpa/2009.75.html>
- Auffhammer, M., Steinhauser, R., 2012. Forecasting The Path of U.S. CO2 Emissions Using State-Level Information. *The Review of Economics and Statistics* 94 (1), 172–185.  
URL [https://doi.org/10.1162/REST\\_a\\_00152](https://doi.org/10.1162/REST_a_00152)
- Azomahou, T., Laisney, F., Van, P. N., 2006. Economic development and CO2 emissions: A nonparametric panel approach. *Journal of Public Economics* 90 (6), 1347 – 1363.  
URL <http://www.sciencedirect.com/science/article/pii/S0047272705001301>
- Barros, F. G., Mendonça, A. F., Nogueira, J. M., 2002. Poverty and environmental degradation: The Kuznets environmental curve for the Brazilian case. Universidade de Brasília (267).
- Bennedsen, M., Hillebrand, E., Jensen, S., 2021. Apocalypse Now? Projecting CO2 Emissions with Neural Networks, Chapter 2 of this dissertation.
- Bertinelli, L., Strobl, E., 2005. The Environmental Kuznets Curve semi-parametrically revisited. *Economics Letters* 88 (3), 350 – 357.  
URL <http://www.sciencedirect.com/science/article/pii/S0165176505001382>
- Borghesi, S., 2006. Income inequality and the environmental Kuznets curve. *Environment, inequality and collective action* 33.
- Cole, M. A., 2003. Development, trade, and the environment: how robust is the Environmental Kuznets Curve? *Environment and Development Economics* 8 (4), 557–580.

- Cole, M. A., 2004. Trade, the pollution haven hypothesis and the environmental Kuznets curve: examining the linkages. *Ecological Economics* 48 (1), 71 – 81.  
 URL <http://www.sciencedirect.com/science/article/pii/S0921800903002556>
- Costantini, V., Martini, C., 2006. A modified environmental Kuznets curve for sustainable development assessment using panel data.
- Cybenko, G., Dec 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* 2 (4), 303–314.  
 URL <https://doi.org/10.1007/BF02551274>
- Dasgupta, S., Hamilton, K., Pandey, K. D., Wheeler, D., 2006. Environment During growth: Accounting for governance and vulnerability. *World Development* 34 (9), 1597 – 1611.  
 URL <http://www.sciencedirect.com/science/article/pii/S0305750X06001070>
- de Bruyn, S., van den Bergh, J., Opschoor, J., 1998. Economic growth and emissions: reconsidering the empirical basis of environmental Kuznets curves. *Ecological Economics* 25 (2), 161 – 175.  
 URL <http://www.sciencedirect.com/science/article/pii/S092180099700178X>
- Dijkgraaf, E., Vollebergh, H. R. J., Oct. 2005. A Test for Parameter Homogeneity in CO<sub>2</sub> Panel EKC Estimations. *Environmental and Resource Economics* 32 (2), 229—239.  
 URL <https://doi.org/10.1007/s10640-005-2776-0>
- Dinda, S., 2004. Environmental Kuznets Curve Hypothesis: A Survey. *Ecological Economics* 49 (4), 431 – 455.  
 URL <http://www.sciencedirect.com/science/article/pii/S0921800904001570>
- Dutt, K., Aug. 2009. Governance, institutions and the environment-income relationship: a cross-country study. *Environment, Development and Sustainability* 11 (4), 705—723.  
 URL <https://doi.org/10.1007/s10668-007-9138-8>
- Galeotti, M., Lanza, A., Pauli, F., 2006. Reassessing the environmental Kuznets curve for CO<sub>2</sub> emissions: A robustness exercise. *Ecological Economics* 57 (1), 152 – 163.  
 URL <http://www.sciencedirect.com/science/article/pii/S0921800905001680>
- Global Carbon Project, 2019. Supplemental data of Global Carbon Budget 2019 (Version 1.0) [Data set]. Global Carbon Project.
- Glorot, X., Bordes, A., Bengio, Y., 11–13 Apr 2011. Deep Sparse Rectifier Neural Networks. In: Gordon, G., Dunson, D., Dudík, M. (Eds.), *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Vol. 15 of *Proceedings of Machine Learning Research*. PMLR, Fort Lauderdale, FL, USA, 315–323.  
 URL <http://proceedings.mlr.press/v15/glorot11a.html>

- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, <http://www.deeplearningbook.org>.
- Grossman, G. M., Krueger, A. B., November 1991. Environmental Impacts of a North American Free Trade Agreement. Working Paper 3914, National Bureau of Economic Research.  
URL <http://www.nber.org/papers/w3914>
- Gu, S., Kelly, B., Xiu, D., 02 2020. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies* 33 (5), 2223–2273.  
URL <https://doi.org/10.1093/rfs/hhaa009>
- Halkos, G. E., Tsionas, E. G., 2001. Environmental Kuznets curves: Bayesian evidence from switching regime models. *Energy Economics* 23 (2), 191 – 210.  
URL <http://www.sciencedirect.com/science/article/pii/S014098830000633>
- Harbaugh, W. T., Levinson, A., Wilson, D. M., 2002. Reexamining the Empirical Evidence for an Environmental Kuznets Curve. *The Review of Economics and Statistics* 84 (3), 541–551.  
URL <http://www.jstor.org/stable/3211570>
- He, K., Zhang, X., Ren, S., Sun, J., December 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: *The IEEE International Conference on Computer Vision (ICCV)*.
- Hendry, D., 2018. First-in, first-out: Driving the UK's per capita carbon dioxide emissions below 1860 levels. VoxEU.
- Hill, R. J., Magnani, E., 2002. An Exploration of the Conceptual and Empirical Basis of the Environmental Kuznets Curve. *Australian Economic Papers* 41 (2), 239–254.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8454.00162>
- Holtz-Eakin, D., Selden, T. M., 1995. Stoking the fires? CO<sub>2</sub> emissions and economic growth. *Journal of Public Economics* 57 (1), 85 – 101.  
URL <http://www.sciencedirect.com/science/article/pii/004727279401449X>
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2 (5), 359–366.  
URL <https://www.sciencedirect.com/science/article/pii/0893608089900208>
- IPCC, 2014. AR5 Synthesis Report: Climate Change 2014.
- Kearsley, A., Riddel, M., 2010. A further inquiry into the Pollution Haven Hypothesis and the Environmental Kuznets Curve. *Ecological Economics* 69 (4), 905 – 919, special Section: Coevolutionary Ecological Economics: Theory and Applications.  
URL <http://www.sciencedirect.com/science/article/pii/S0921800909004637>

- Kingma, D. P., Ba, J., 2014. Adam: A Method for Stochastic Optimization.
- Kuznets, S., 1955. Economic Growth and Income Inequality. *The American Economic Review* 45 (1), 1–28.  
URL <http://www.jstor.org/stable/1811581>
- Lee, C.-C., Lee, J.-D., 2009. Income and CO<sub>2</sub> emissions: Evidence from panel unit root and cointegration tests. *Energy Policy* 37 (2), 413 – 423.  
URL <http://www.sciencedirect.com/science/article/pii/S0301421508005181>
- Leshno, M., Lin, V. Y., Pinkus, A., Schocken, S., 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks* 6 (6), 861–867.  
URL <https://www.sciencedirect.com/science/article/pii/S0893608005801315>
- Levin, A., Lin, C.-F., Chu, C.-S. J., 2002. Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics* 108 (1), 1–24.  
URL <https://www.sciencedirect.com/science/article/pii/S0304407601000987>
- List, J. A., Gallet, C. A., 1999. The environmental Kuznets curve: does one size fit all? *Ecological Economics* 31 (3), 409 – 423.  
URL <http://www.sciencedirect.com/science/article/pii/S0921800999000646>
- Magnani, E., 2001. The Environmental Kuznets Curve: development path or policy result? *Environmental Modelling & Software* 16 (2), 157 – 165, environmental Modelling and Socioeconomics.  
URL <http://www.sciencedirect.com/science/article/pii/S1364815200000797>
- Malec, P., Schienle, M., 2014. Nonparametric kernel density estimation near the boundary. *Computational Statistics & Data Analysis* 72, 57 – 76.  
URL <http://www.sciencedirect.com/science/article/pii/S0167947313003770>
- Masters, T., 1993. Practical neural network recipes in C++. (Academic Press).
- Millimet, D. L., List, J. A., Stengos, T., 2003. The Environmental Kuznets Curve: Real Progress or Misspecified Models? *The Review of Economics and Statistics* 85 (4), 1038–1047.  
URL <http://www.jstor.org/stable/3211824>
- Mundlak, Y., 1978. On the Pooling of Time Series and Cross Section Data. *Econometrica* 46 (1), 69–85.  
URL <http://www.jstor.org/stable/1913646>
- Müller-Fürstenberger, G., Wagner, M., 2007. Exploring the environmental Kuznets hypothesis: Theoretical and econometric problems. *Ecological Economics* 62 (3), 648 – 660.  
URL <http://www.sciencedirect.com/science/article/pii/S0921800906004058>

- NOAA, 2020. The NOAA Annual Greenhouse Gas Index. Accessed June 2020.  
<https://www.esrl.noaa.gov/gmd/aggi/aggi.html>  
URL <https://www.esrl.noaa.gov/gmd/aggi/aggi.html>
- Panayotou, T., 1997. Demystifying the environmental Kuznets curve: turning a black box into a policy tool. *Environment and Development Economics* 2 (4), 465–484.
- Perman, R., Stern, D., 1999. The Environmental Kuznets Curve: Implications of Non-Stationarity. Working papers in ecological economics, Australian National University, Centre for Resource and Environmental Studies, Ecological Economics Program.  
URL <https://EconPapers.repec.org/RePEc:anu:wpeep:9901>
- Perman, R., Stern, D. I., 2003. Evidence from panel unit root and cointegration tests that the Environmental Kuznets Curve does not exist. *Australian Journal of Agricultural and Resource Economics* 47 (3), 325–347.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8489.00216>
- Peters, G. P., Hertwich, E. G., Mar. 2008a. CO<sub>2</sub> Embodied in International Trade with Implications for Global Climate Policy. *Environmental Science & Technology* 42 (5), 1401–1407.  
URL <https://doi.org/10.1021/es072023k>
- Peters, G. P., Hertwich, E. G., Jan. 2008b. Post-Kyoto greenhouse gas inventories: production versus consumption. *Climatic Change* 86 (1), 51–66.  
URL <https://doi.org/10.1007/s10584-007-9280-1>
- Peters, G. P., Minx, J. C., Weber, C. L., Edelenhofer, O., 2011. Growth in emission transfers via international trade from 1990 to 2008. *Proceedings of the National Academy of Sciences* 108 (21), 8903–8908.  
URL <https://www.pnas.org/content/108/21/8903>
- Ramachandran, P., Zoph, B., Le, Q. V., 2017. Searching for Activation Functions.
- Riahi et al., 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change* 42, 153 – 168.  
URL <http://www.sciencedirect.com/science/article/pii/S0959378016300681>
- Richmond, A. K., Kaufmann, R. K., February 2006. Is there a turning point in the relationship between income and energy use and/or carbon emissions? *Ecological Economics* 56 (2), 176–189.  
URL <https://ideas.repec.org/a/eee/ecolec/v56y2006i2p176-189.html>
- Romero-Ávila, D., 2008. Questioning the empirical basis of the environmental Kuznets curve for CO<sub>2</sub>: New evidence from a panel stationarity test robust to multiple

- breaks and cross-dependence. *Ecological Economics* 64 (3), 559 – 574.  
 URL <http://www.sciencedirect.com/science/article/pii/S0921800907002042>
- Schmalensee, R., Stoker, T., Judson, R., 02 1998. World Carbon Dioxide Emissions: 1950-2050. *The Review of Economics and Statistics* 80, 15–27.
- Selden, T. M., Song, D., 1994. Environmental Quality and Development: Is There a Kuznets Curve for Air Pollution Emissions? *Journal of Environmental Economics and Management* 27 (2), 147 – 162.  
 URL <http://www.sciencedirect.com/science/article/pii/S009506968471031X>
- Shafik, N., 1994. Economic Development and Environmental Quality: An Econometric Analysis. *Oxford Economic Papers* 46, 757–773.  
 URL <http://www.jstor.org/stable/2663498>
- Stern, D., Common, M., 02 2001. Is There An Environmental Kuznets Curve for Sulphur. *Journal of Environmental Economics and Management* 41, 162–178.
- Stern, D. I., 2004. The Rise and Fall of the Environmental Kuznets Curve. *World Development* 32 (8), 1419 – 1439.  
 URL <http://www.sciencedirect.com/science/article/pii/S0305750X04000798>
- Stern, D. I., 2010. Between estimates of the emissions-income elasticity. *Ecological Economics* 69 (11), 2173 – 2182, special Section - Payments for Ecosystem Services: From Local to Global.  
 URL <http://www.sciencedirect.com/science/article/pii/S092180091000265X>
- Stern, D. I., Apr 2017. The environmental Kuznets curve after 25 years. *Journal of Bioeconomics* 19 (1), 7–28.  
 URL <https://doi.org/10.1007/s10818-017-9243-1>
- Stern, D. I., Common, M. S., Barbier, E. B., 1996. Economic growth and environmental degradation: The environmental Kuznets curve and sustainable development. *World Development* 24 (7), 1151 – 1160.  
 URL <http://www.sciencedirect.com/science/article/pii/0305750X96000320>
- Suri, V., Chapman, D., 1998. Economic growth, trade and energy: implications for the environmental Kuznets curve. *Ecological Economics* 25 (2), 195 – 208.  
 URL <http://www.sciencedirect.com/science/article/pii/S0921800997001808>
- Taskin, F., Zaim, O., 2000. Searching for a Kuznets curve in environmental efficiency using kernel estimation. *Economics Letters* 68 (2), 217 – 223.  
 URL <http://www.sciencedirect.com/science/article/pii/S0165176500002500>
- Torras, M., Boyce, J. K., 1998. Income, inequality, and pollution: a reassessment of the environmental Kuznets Curve. *Ecological Economics* 25 (2), 147 – 160.  
 URL <http://www.sciencedirect.com/science/article/pii/S0921800997001778>

Tsurumi, T., Managi, S., 2015. Environmental Kuznets curve: Economic growth and emission reduction. Cited By 3.

URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84943553413&doi=10.4324%2f9781315745800&partnerID=40&md5=65a4dd7392df658a6e0cc2fa45677c83>

Vollebergh, H. R., Melenberg, B., Dijkgraaf, E., 2009. Identifying reduced-form relations with panel data: The case of pollution and income. Journal of Environmental Economics and Management 58 (1), 27 – 42.

URL <http://www.sciencedirect.com/science/article/pii/S0095069609000187>

Wagner, M., 2008. The carbon Kuznets curve: A cloudy picture emitted by bad econometrics? Resource and Energy Economics 30 (3), 388 – 408.

URL <http://www.sciencedirect.com/science/article/pii/S0928765507000504>

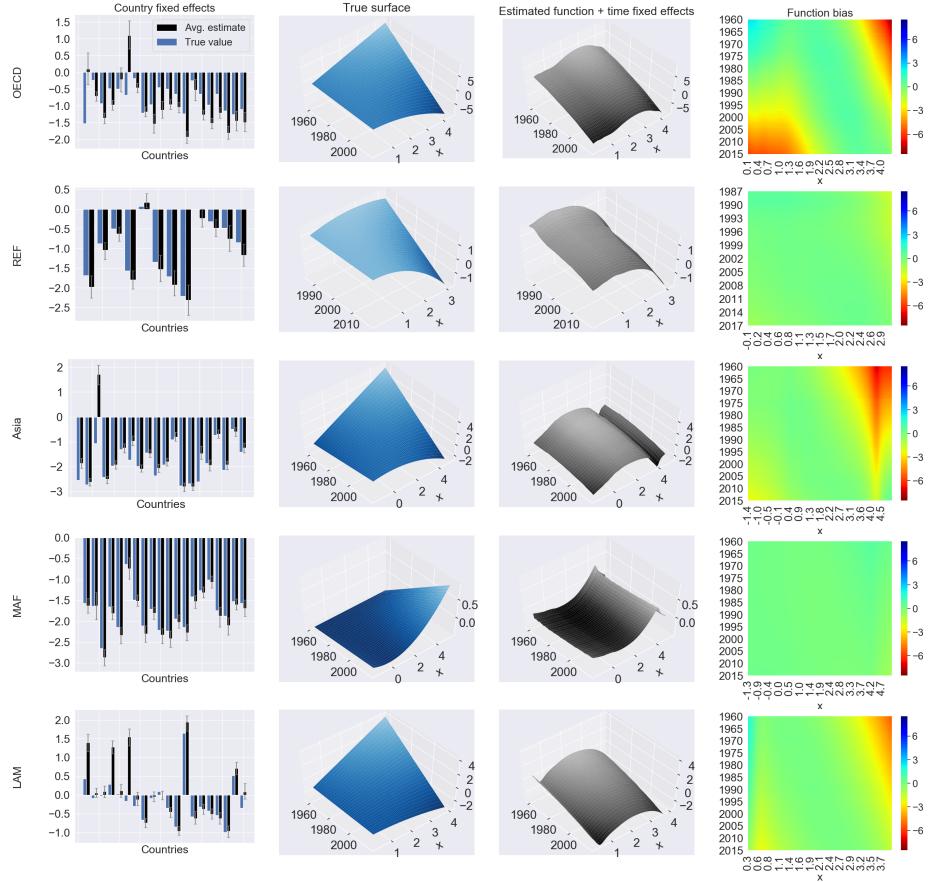
Wagner, M., 2015. The Environmental Kuznets Curve, Cointegration and Nonlinearity. Journal of Applied Econometrics 30 (6), 948–967.

URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2421>

## Appendix

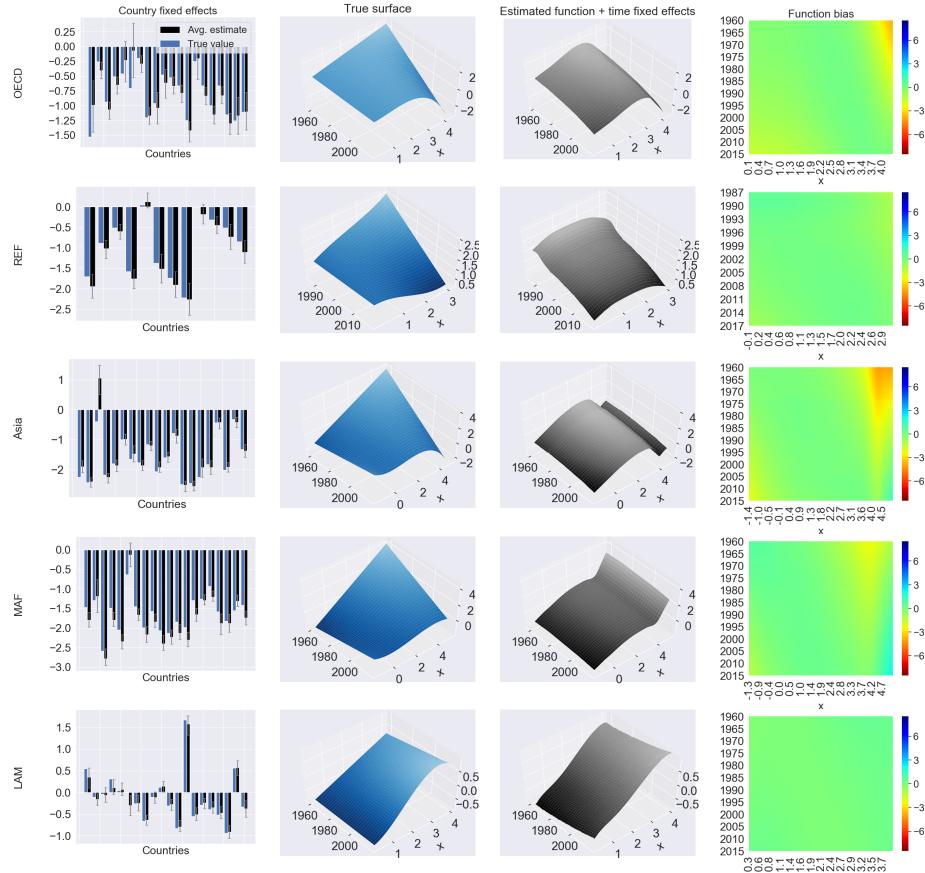
### A.1 Figures

**Figure A.1:** Monte Carlo misspecification results for the regional formulation of the static model: linear-quadratic function

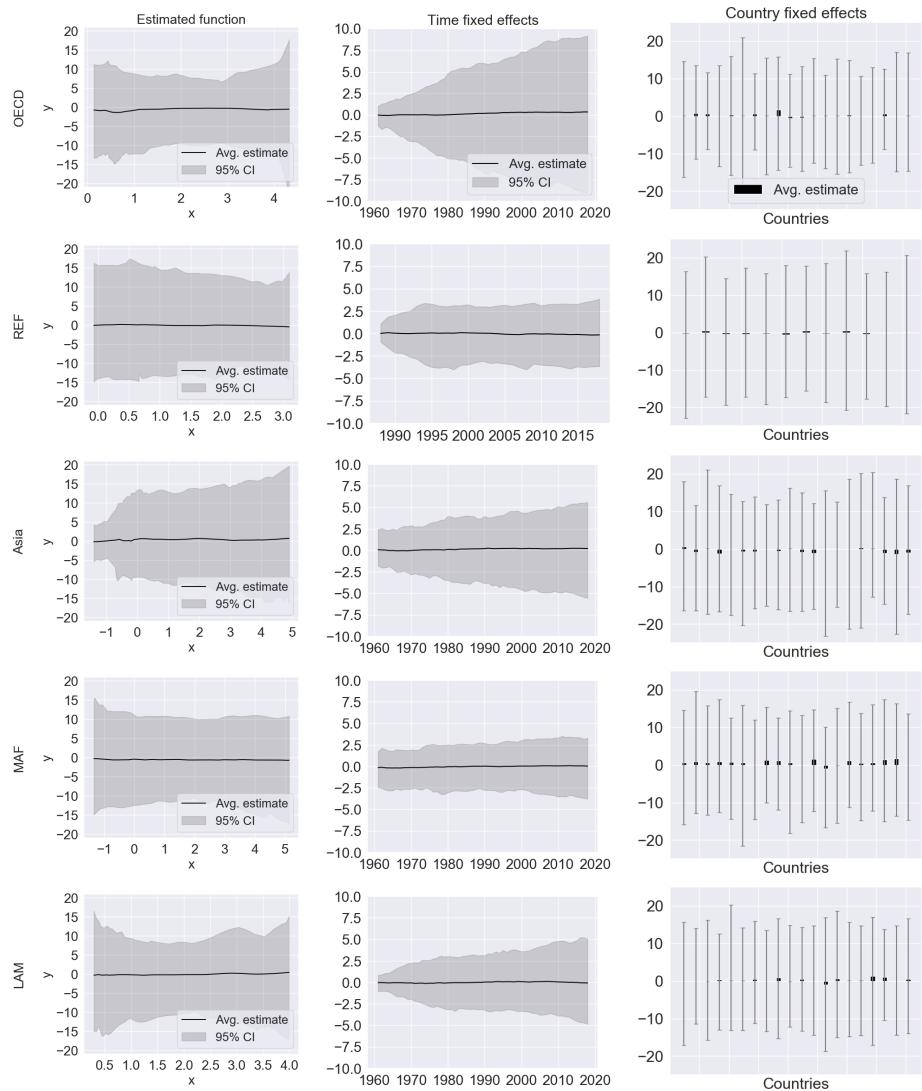


Note: In the country fixed effects plots, gray error bars indicate 95% confidence intervals.

**Figure A.2:** Monte Carlo misspecification results for the regional formulation of the static model: linear-cubic function

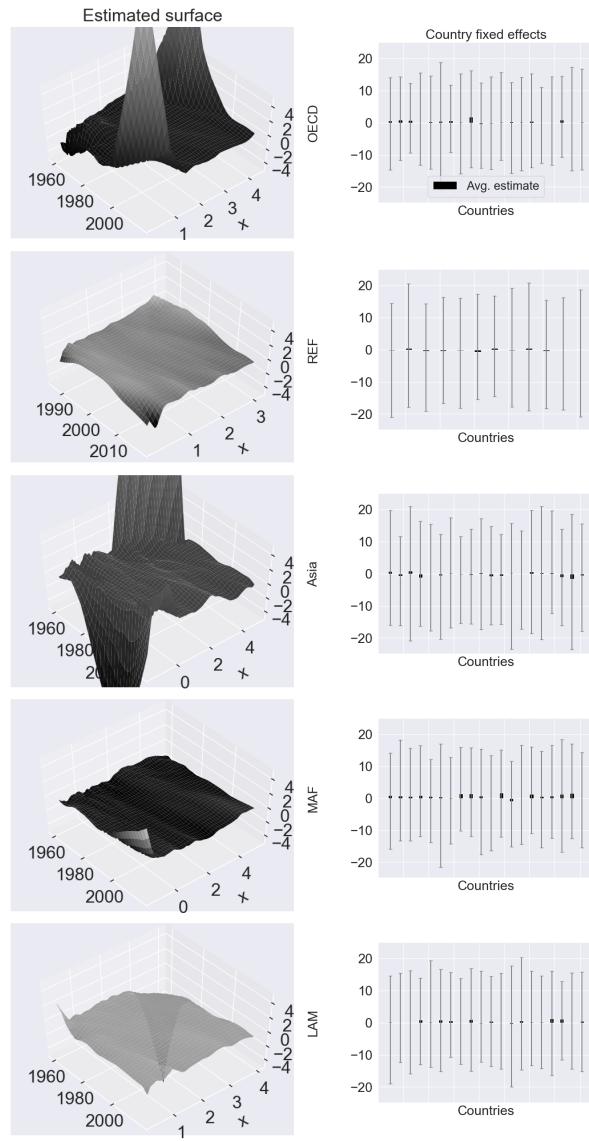


Note: In the country fixed effects plots, gray error bars indicate 95% confidence intervals.

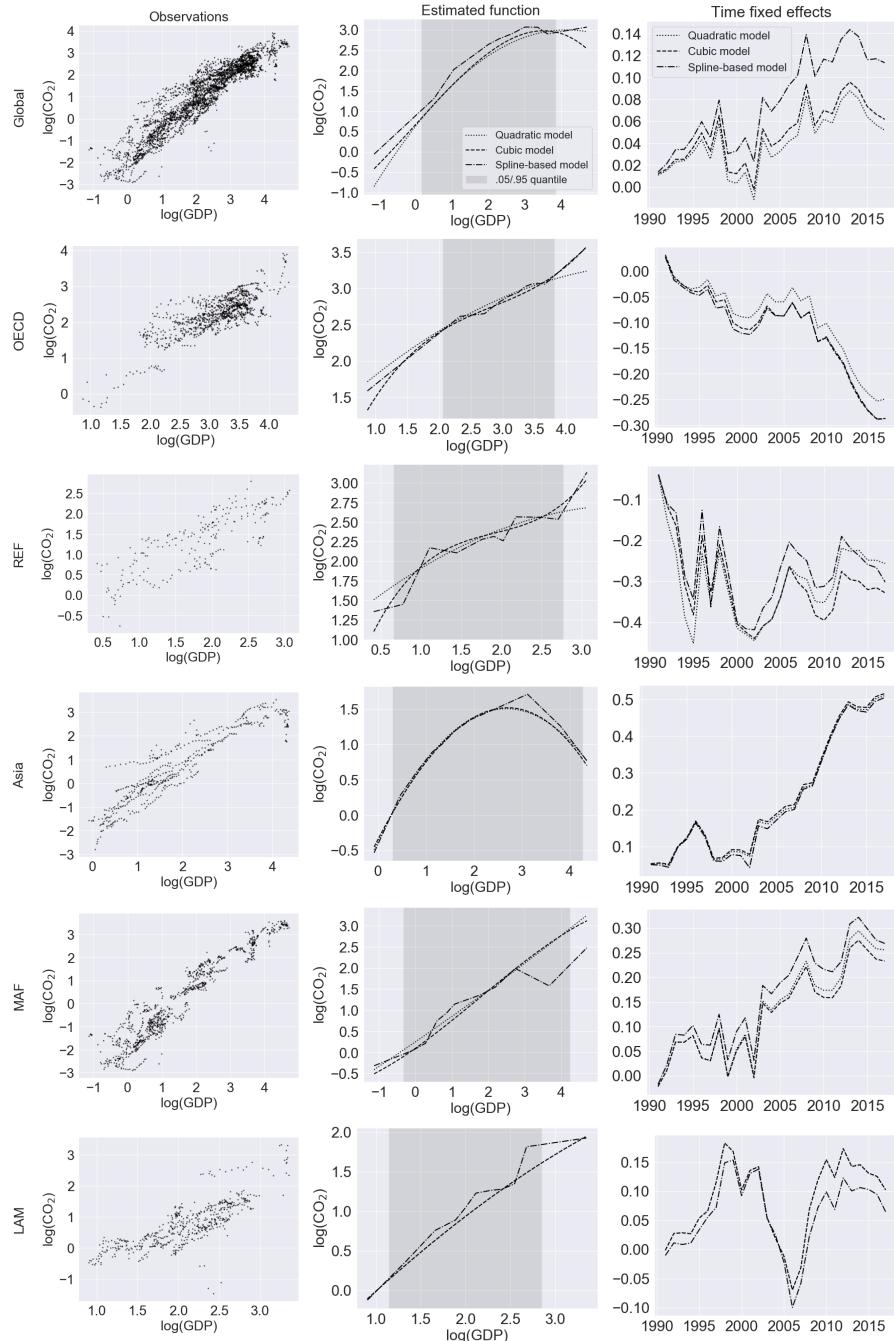
**Figure A.3:** Monte Carlo spuriousness results for the regional formulation of the static model

Note: In the country fixed effects plots, gray error bars indicate 95% confidence intervals.

**Figure A.4:** Monte Carlo spuriousness results for the regional formulation of the dynamic model

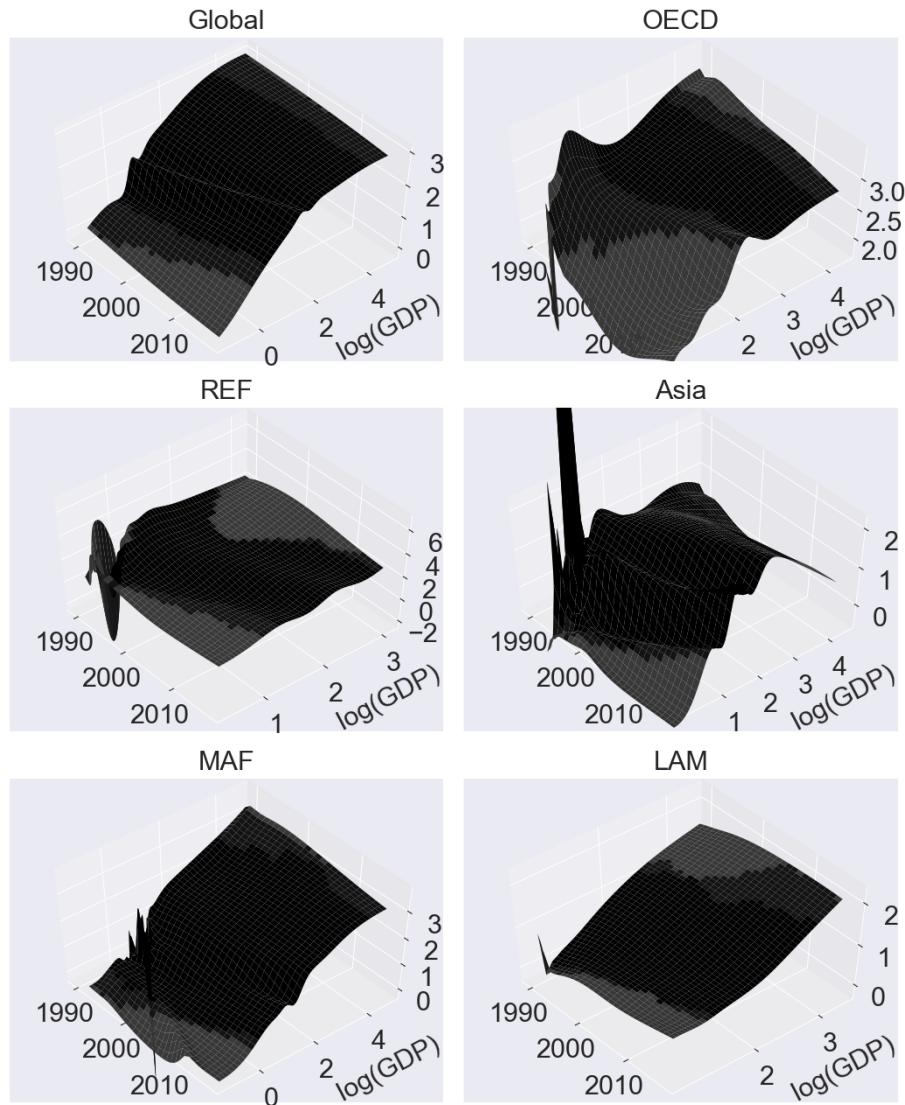


Note: In the country fixed effects plots, gray error bars indicate 95% confidence intervals.

**Figure A.5:** Estimation results for the benchmark models and consumption-based emissions

Note: Estimation results in the first row of plots are based on global model formulations where every country is used for estimation. Remaining rows show estimation results from region-wise estimation of the benchmark models.

**Figure A.6:** Estimation results for the dynamic model and territorial emissions based on a restricted sample that coincides with the one for consumption-based emissions



Note: Black coloring indicates area between the .05 and .95 income quantiles. Dark gray coloring indicates area with data support. Light grey coloring indicates area with no data support. The estimated surface in the top left panel is based on the global formulation of the dynamic neural network model. Remaining surfaces are based on the regional formulation of the dynamic neural network model.

## A.2 Tables

**Table A.1:** Macro-region definitions

<b>OECD</b>		
Albania <sup>1,2</sup>	Australia <sup>1,2</sup>	Austria <sup>1,2</sup>
Belgium <sup>1,2</sup>	Bosnia and Herzegovina <sup>1</sup>	Bulgaria <sup>1,2</sup>
Canada <sup>1,2</sup>	Croatia <sup>1,2</sup>	Cyprus <sup>1,2</sup>
Czech Republic <sup>1,2</sup>	Denmark <sup>1,2</sup>	Estonia <sup>1,2</sup>
Finland <sup>1,2</sup>	France <sup>1,2</sup>	Germany <sup>1,2</sup>
Greece <sup>1,2</sup>	Hungary <sup>1,2</sup>	Iceland <sup>1</sup>
Ireland <sup>1,2</sup>	Italy <sup>1,2</sup>	Japan <sup>1,2</sup>
Latvia <sup>1,2</sup>	Lithuania <sup>1,2</sup>	Luxembourg <sup>1,2</sup>
Malta <sup>1,2</sup>	Montenegro <sup>1</sup>	Netherlands <sup>1,2</sup>
New Zealand <sup>1,2</sup>	Norway <sup>1,2</sup>	Poland <sup>1,2</sup>
Portugal <sup>1,2</sup>	Puerto Rico	Romania <sup>1,2</sup>
<b>Serbia</b>	Slovakia <sup>1,2</sup>	Slovenia <sup>1,2</sup>
Spain <sup>1,2</sup>	Sweden <sup>1,2</sup>	Switzerland <sup>1,2</sup>
North Macedonia <sup>1</sup>	Turkey <sup>1,2</sup>	United Kingdom <sup>1,2</sup>
United States <sup>1,2</sup>		
<b>REF</b>		
Armenia <sup>1,2</sup>	Azerbaijan <sup>1,2</sup>	Belarus <sup>1,2</sup>
Georgia <sup>1,2</sup>	Kazakhstan <sup>1,2</sup>	Kyrgyzstan <sup>1,2</sup>
Moldova <sup>1</sup>	Russia <sup>1,2</sup>	Tajikistan <sup>1</sup>
Turkmenistan <sup>1</sup>	Ukraine <sup>1,2</sup>	Uzbekistan <sup>1</sup>
<b>Kosovo<sup>1</sup></b>		
<b>Asia</b>		
Afghanistan <sup>1</sup>	Bangladesh <sup>1,2</sup>	Bhutan <sup>1</sup>
Brunei Darussalam <sup>1,2</sup>	Cambodia <sup>1,2</sup>	China <sup>1,2</sup>
South Korea <sup>1,2</sup>	Fiji <sup>1</sup>	French Polynesia
India <sup>1,2</sup>	Indonesia <sup>1,2</sup>	Laos <sup>1,2</sup>
Malaysia <sup>1,2</sup>	Maldives <sup>1</sup>	Micronesia <sup>1</sup>
Mongolia <sup>1,2</sup>	Myanmar <sup>1</sup>	Nepal <sup>1,2</sup>
<b>New Caledonia</b>	Pakistan <sup>1,2</sup>	Papua New Guinea <sup>1</sup>
Philippines <sup>1,2</sup>	Samoa <sup>1</sup>	Singapore <sup>1,2</sup>
Solomon Islands <sup>1</sup>	Sri Lanka <sup>1,2</sup>	Taiwan
Thailand <sup>1,2</sup>	Timor-Leste <sup>1</sup>	Vanuatu <sup>1</sup>
Vietnam <sup>1,2</sup>	<b>Tuvalu<sup>1</sup></b>	Macao <sup>1</sup>
<b>Marshall Islands<sup>1</sup></b>	Palau <sup>1</sup>	Hong Kong <sup>1,2</sup>
Tonga <sup>1</sup>	Kiribati <sup>1</sup>	
<b>MAF</b>		
Algeria <sup>1</sup>	Angola <sup>1</sup>	Bahrain <sup>1,2</sup>
Benin <sup>1,2</sup>	Botswana <sup>1,2</sup>	Burkina Faso <sup>1,2</sup>

Burundi <sup>1</sup>	Cameroon <sup>1,2</sup>	Cabo Verde <sup>1</sup>
Central African Republic <sup>1</sup>	Chad <sup>1</sup>	Comoros <sup>1</sup>
Congo (DRC) <sup>1</sup>	Congo (RDC) <sup>1</sup>	Côte d'Ivoire <sup>1,2</sup>
<b>Djibouti</b>	Egypt <sup>1,2</sup>	Equatorial Guinea <sup>1</sup>
Eritrea <sup>1</sup>	Ethiopia <sup>1,2</sup>	Gabon <sup>1</sup>
Gambia <sup>1</sup>	Ghana <sup>1,2</sup>	Guinea <sup>1,2</sup>
Guinea-Bissau <sup>1</sup>	Iran <sup>1,2</sup>	Iraq <sup>1</sup>
Israel <sup>1,2</sup>	Jordan <sup>1,2</sup>	Kenya <sup>1,2</sup>
Kuwait <sup>1,2</sup>	Lebanon <sup>1</sup>	Lesotho <sup>1</sup>
Liberia <sup>1</sup>	Libya <sup>1</sup>	<b>Madagascar</b>
Malawi <sup>1,2</sup>	Mali <sup>1,2</sup>	Mauritania <sup>1</sup>
Mauritius <sup>1,2</sup>	Morocco <sup>1,2</sup>	Mozambique <sup>1,2</sup>
Namibia <sup>1,2</sup>	Niger <sup>1</sup>	Nigeria <sup>1,2</sup>
Palestine <sup>1</sup>	Oman <sup>1,2</sup>	Qatar <sup>1,2</sup>
Rwanda <sup>1,2</sup>	Saudi Arabia <sup>1,2</sup>	Senegal <sup>1,2</sup>
Sierra Leone <sup>1</sup>	<b>Somalia</b>	South Africa <sup>1,2</sup>
Sudan <sup>1</sup>	Eswatini <sup>1</sup>	<b>Syria</b>
Togo <sup>1,2</sup>	Tunisia <sup>1,2</sup>	Uganda <sup>1,2</sup>
United Arab Emirates <sup>1,2</sup>	Tanzania <sup>1,2</sup>	Yemen <sup>1</sup>
Zambia <sup>1,2</sup>	Zimbabwe <sup>1,2</sup>	Seychelles <sup>1</sup>
Sao Tome and Principe <sup>1</sup>		
<hr/>		
<b>LAM</b>		
Argentina <sup>1,2</sup>	Aruba <sup>1</sup>	Bahamas <sup>1</sup>
Barbados <sup>1</sup>	Belize <sup>1</sup>	Bolivia <sup>1,2</sup>
Brazil <sup>1,2</sup>	Chile <sup>1,2</sup>	Colombia <sup>1,2</sup>
Costa Rica <sup>1,2</sup>	<b>Cuba</b>	Dominican Republic <sup>1,2</sup>
Ecudao <sup>1,2</sup>	El Salvador <sup>1,2</sup>	<b>Grenada</b> <sup>1</sup>
Guatemala <sup>1,2</sup>	Guyana <sup>1</sup>	Haiti <sup>1</sup>
Honduras <sup>1,2</sup>	Jamaica <sup>1,2</sup>	Mexico <sup>1,2</sup>
Nicaragua <sup>1,2</sup>	Panama <sup>1,2</sup>	Paraguay <sup>1,2</sup>
Peru <sup>1,2</sup>	Suriname <sup>1</sup>	Trinidad and Tobago <sup>1,2</sup>
Uruguay <sup>1,2</sup>	Venezuela <sup>1,2</sup>	Saint Vincent <sup>1</sup>
<b>Curaçao</b> <sup>1</sup>	Saint Lucia <sup>1</sup>	<b>Antigua and Barbuda</b> <sup>1</sup>
<b>Dominica</b> <sup>1</sup>	Bermuda <sup>1</sup>	
<hr/>		

Note: <sup>1</sup>Country is in the CO<sub>2</sub> panel. <sup>2</sup>Country is in the CO<sub>2</sub><sup>C</sup> and CO<sub>2</sub><sup>★</sup> panel. Region definitions are from the Shared Socioeconomic Pathways (SSPs; Riahi et al. [2017]): "OECD" is OECD90 and EU member states and candidates; "REF" is reforming economies Eastern Europe and the former Soviet Union; "Asia" is Asian countries excluding the Middle East, Japan and former Soviet Union states; "MAF" is the Middle East and Africa; and "LAM" is Latin America and the Caribbean. Countries in red are in the SSP database but not in our data set. Countries in green are in our data set but not in the SSP database.

**Table A.2:** Descriptive statistics: mean

	Mean						
	1960	1970	1980	1990	2000	2010	2017
<b>Global</b>							
GDP	3.70	5.49	6.78	7.69	9.01	11.46	13.50
CO <sub>2</sub>	2.51	3.51	3.68	4.19	4.01	4.66	4.64
CO <sub>2</sub> *	2.58	3.62	3.80	4.34	4.17	4.90	4.90
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	4.35	4.19	4.89	4.89
<b>OECD</b>							
GDP	10.84	16.09	20.47	24.76	29.45	32.62	35.99
CO <sub>2</sub>	8.22	12.08	12.38	12.24	12.31	11.22	10.18
CO <sub>2</sub> *	8.22	12.08	12.39	12.25	12.36	11.25	10.21
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	13.02	13.52	12.69	11.51
<b>REF</b>							
GDP	NA	NA	NA	10.29	6.52	11.16	12.30
CO <sub>2</sub>	NA	NA	NA	14.06	7.69	8.62	8.44
CO <sub>2</sub> *	NA	NA	NA	15.09	8.26	9.57	9.51
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	12.59	6.60	7.87	8.29
<b>Asia</b>							
GDP	0.78	1.00	1.37	2.14	3.52	6.54	9.51
CO <sub>2</sub>	0.68	0.65	1.00	1.39	1.79	3.37	3.82
CO <sub>2</sub> *	0.69	0.66	1.02	1.41	1.82	3.45	3.91
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	1.44	1.74	3.13	3.59
<b>MAF</b>							
GDP	3.23	5.61	6.84	5.06	5.55	6.86	7.21
CO <sub>2</sub>	0.89	1.44	2.09	1.69	2.11	2.46	2.46
CO <sub>2</sub> *	1.10	1.71	2.41	1.91	2.45	2.88	2.87
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	1.68	1.90	2.72	2.73
<b>LAM</b>							
GDP	5.03	6.54	9.13	8.53	9.84	11.86	12.44
CO <sub>2</sub>	1.32	1.76	2.40	2.24	2.58	2.88	2.73
CO <sub>2</sub> *	1.34	1.78	2.42	2.27	2.60	2.92	2.76
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	2.26	2.49	3.05	3.00

Note: "CO<sub>2</sub><sup>C</sup>" are consumption-based CO<sub>2</sub> emissions; "CO<sub>2</sub>\*" are CO<sub>2</sub> emissions based on a restricted sample that coincides with the one for consumption-based emissions; and GDP is from the CO<sub>2</sub> panel. For given region  $r$  and time period  $t$ , mean values are calculated as the per capita values for that region: Mean  $\equiv \sum_{i \in I_r} \frac{x_{it}}{\text{POP}_{it}} \frac{\text{POP}_{it}}{\text{POP}_t^r} = \sum_{i \in I_r} \frac{x_{it}}{\text{POP}_t^r}$ , where  $I_r \subseteq \{1, 2, \dots, N\}$  is the set of indices of countries belonging to region  $r$ ,  $N$  is the total number of countries, POP is population size,  $\text{POP}_t^r \equiv \sum_{i \in I_r} \text{POP}_{it}$ , and  $x \in \{\text{GDP}, \text{CO}_2, \text{CO}_2^*, \text{CO}_2^C\}$ .

**Table A.3:** Descriptive statistics: standard deviation

	Standard deviation						
	1960	1970	1980	1990	2000	2010	2017
<b>Global</b>							
GDP	4.71	7.24	9.43	9.79	11.35	11.67	12.39
CO <sub>2</sub>	4.37	5.81	5.58	5.65	5.20	4.75	4.42
CO <sub>2</sub> *	4.43	5.89	5.66	5.74	5.29	4.80	4.45
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	5.72	5.55	5.00	4.59
<b>OECD</b>							
GDP	4.16	4.09	5.33	7.88	10.21	10.03	10.51
CO <sub>2</sub>	6.13	6.60	6.22	5.79	6.23	5.29	4.63
CO <sub>2</sub> *	6.13	6.60	6.22	5.80	6.22	5.28	4.63
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	5.30	6.13	5.19	4.62
<b>REF</b>							
GDP	NA	NA	NA	3.26	2.62	4.62	5.14
CO <sub>2</sub>	NA	NA	NA	4.36	2.93	3.86	4.30
CO <sub>2</sub> *	NA	NA	NA	3.26	2.61	3.26	3.70
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	4.92	2.24	2.29	2.64
<b>Asia</b>							
GDP	0.51	0.78	1.74	2.28	3.25	4.78	6.11
CO <sub>2</sub>	0.46	0.45	0.82	1.08	1.41	2.69	2.85
CO <sub>2</sub> *	0.46	0.45	0.82	1.07	1.40	2.68	2.83
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	1.37	1.80	2.47	2.64
<b>MAF</b>							
GDP	2.23	9.60	14.06	7.95	9.16	9.77	10.25
CO <sub>2</sub>	1.55	2.55	3.63	2.98	4.08	4.48	4.53
CO <sub>2</sub> *	1.71	2.87	4.07	3.31	4.58	5.04	5.11
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	2.84	3.27	4.65	4.79
<b>LAM</b>							
GDP	2.46	2.89	2.90	2.55	2.99	3.36	3.27
CO <sub>2</sub>	1.33	1.44	1.82	1.56	1.75	2.21	1.85
CO <sub>2</sub> *	1.33	1.39	1.58	1.54	1.62	2.15	1.77
CO <sub>2</sub> <sup>C</sup>	NA	NA	NA	1.22	1.20	1.45	1.58

Note: "CO<sub>2</sub><sup>C</sup>" are consumption-based CO<sub>2</sub> emissions; "CO<sub>2</sub>\*" are CO<sub>2</sub> emissions based on a restricted sample that coincides with the one for consumption-based emissions; and GDP is from the CO<sub>2</sub> panel. For given region  $r$  and time period  $t$ , standard deviations are calculated using the following relation: Standard deviation  $\equiv \sqrt{\sum_{i \in I_r} \left( \frac{x_{it}}{POP_{it}} - \text{Mean} \right)^2 \frac{POP_{it}}{POP_t^r}}$ , where Mean is from Table A.2.  $I_r \subseteq \{1, 2, \dots, N\}$  is the set of indices of countries belonging to region  $r$ ,  $N$  is the total number of countries, POP is population size,  $POP_t^r \equiv \sum_{i \in I_r} POP_{it}$ , and  $x \in \{GDP, CO_2, CO_2^*, CO_2^C\}$ .

**Table A.4:** Neural network architectures

Network architecture	# parameters in static model			# parameters in dynamic model		
	Global (R = 1)	Regional (R = 5)	National (R = 186)	Global (R = 1)	Regional (R = 5)	National (R = 186)
(2)	6	14	376	8	16	378
(4)	12	28	752	16	32	756
(8)	24	56	1504	32	64	1512
(16)	48	112	3008	64	128	3024
(32)	96	224	6016	128	256	6048
(2,2)	12	20	382	14	22	384
(4,2)	20	28	390	24	32	394
(4,4)	32	48	772	36	52	776
(8,2)	36	44	406	44	52	414
(8,4)	56	72	796	64	80	804
(8,8)	96	128	1576	104	136	1584
(16,2)	68	76	438	84	92	454
(16,4)	104	120	844	120	136	860
(16,8)	176	208	1656	192	224	1672
(16,16)	320	384	3280	336	400	3296
(32,2)	132	140	502	164	172	534
(32,4)	200	216	940	232	248	972
(32,8)	336	368	1816	368	400	1848
(32,16)	608	672	3568	640	704	3600
(32,32)	1152	1280	7072	1184	1312	7104
(2,2,2)	18	26	388	20	28	390
(4,2,2)	26	34	396	30	38	400
(4,4,2)	40	48	410	44	52	414
(4,4,4)	52	68	792	56	72	796
(8,2,2)	42	50	412	50	58	420
(8,4,2)	64	72	434	72	80	442
(8,4,4)	76	92	816	84	100	824
(8,8,2)	108	116	478	116	124	486
(8,8,4)	128	144	868	136	152	876
(8,8,8)	168	200	1648	176	208	1656
(16,2,2)	74	82	444	90	98	460
(16,4,2)	112	120	482	128	136	498
(16,4,4)	124	140	864	140	156	880
(16,8,2)	188	196	558	204	212	574
(16,8,4)	208	224	948	224	240	964
(16,8,8)	248	280	1728	264	296	1744
(16,16,2)	340	348	710	356	364	726
(16,16,4)	376	392	1116	392	408	1132
(16,16,8)	448	480	1928	464	496	1944

(16,16,16)	592	656	3552	608	672	3568
(32,2,2)	138	146	508	170	178	540
(32,4,2)	208	216	578	240	248	610
(32,4,4)	220	236	960	252	268	992
(32,8,2)	348	356	718	380	388	750
(32,8,4)	368	384	1108	400	416	1140
(32,8,8)	408	440	1888	440	472	1920
(32,16,2)	628	636	998	660	668	1030
(32,16,4)	664	680	1404	696	712	1436
(32,16,8)	736	768	2216	768	800	2248
(32,16,16)	880	944	3840	912	976	3872
(32,32,2)	1188	1196	1558	1220	1228	1590
(32,32,4)	1256	1272	1996	1288	1304	2028
(32,32,8)	1392	1424	2872	1424	1456	2904
(32,32,16)	1664	1728	4624	1696	1760	4656
(32,32,32)	2208	2336	8128	2240	2368	8160

Note: " $(a,b,c)$ " indicates a neural network architecture with three hidden layers containing  $a$  units in the first layer,  $b$  in the second, and  $c$  in the third; "# parameters" is the number of free model parameters excluding fixed effects.

**Table A.5:** Adam algorithm

---

**Require:** Step size  $\epsilon$  (0.001)  
**Require:** Exponential decay rates  $\rho_1, \rho_2 \in [0, 1]$  for moment estimates (0.9, 0.999)  
**Require:** Small constant  $\delta$  used for numerical stabilization ( $10^{-8}$ )  
**Require:** Initial parameters  $\theta$

Initialize first moment vector:  $m = 0$   
Initialize second moment vector:  $v = 0$   
Initialize time step:  $t = 0$

**while** stopping criterion not met **do**

- $t \leftarrow t + 1$
- Compute gradient:  $g \leftarrow \nabla_{\theta} J(\theta)$
- Update biased first moment estimate:  $m \leftarrow \rho_1 m + (1 - \rho_1)g$
- Update biased second moment estimate:  $v \leftarrow \rho_2 v + (1 - \rho_2)g \odot g$
- Correct bias in first moment:  $\hat{m} \leftarrow m / (1 - \rho_1^t)$
- Correct bias in second moment:  $\hat{v} \leftarrow v / (1 - \rho_2^t)$
- compute update:  $\Delta\theta = -\epsilon \cdot \hat{m} / (\sqrt{\hat{v}} + \delta)$
- Apply update:  $\theta \leftarrow \theta + \Delta\theta$

**end while**

---

Note: Numbers in parentheses are suggested defaults.

The Adam algorithm individually adapts the learning rate of all parameters in two ways. First, its scales them inversely proportional to the square root of an exponentially decaying average of past squared values of the gradient (second moment estimate). In this way, the learning rate of parameters with small partial derivatives of the loss function decreases less rapidly than that of parameters with large partial derivatives. This implies greater progress in more gently sloped regions of the parameter space. Second, to speed up optimization, especially in face of pathological curvature, Adam incorporates momentum by scaling the individual learning rates proportionally to an exponentially decaying average of past values of the gradient (first moment estimate). Finally, Adam includes bias corrections of the moment estimates to account for initialization at the origin.

**Table A.6:** Optimal neural network architectures for the static model in the Monte Carlo misspecification experiment

	Function	
	Linear-quadratic	Linear-cubic
<b>Global model formulation</b>		
Neural network architecture	(2,2)	(2,2,2)
# parameters (excl. fixed effects)	12	18
<b>Regional model formulation</b>		
Neural network architecture	(4,4,4)	(4,4)
# parameters (excl. fixed effects)	52	32

Note: " $(a,b,c)$ " indicates a neural network architecture with three hidden layers containing  $a$  units in the first layer,  $b$  in the second, and  $c$  in the third.

**Table A.7:** Optimal neural network architectures for the Monte Carlo spuriousness experiment

	Formulation	
	Global	Regional
Static model specification		
Neural network architecture	(4,2)	(4,4)
# parameters (excl. fixed effects)	20	48
Dynamic model specification		
Neural network architecture	(4,2)	(16,8)
# parameters (excl. fixed effects)	24	224

Note: " $(a,b,c)$ " indicates a neural network architecture with three hidden layers containing  $a$  units in the first layer,  $b$  in the second, and  $c$  in the third.

# 2

CHAPTER

## APOCALYPSE NOW? PROJECTING CO<sub>2</sub> EMISSIONS WITH NEURAL NETWORKS

**Mikkel Bennedsen**

*Aarhus University and CREATEs*

**Eric Hillebrand**

*Aarhus University and CREATEs*

**Sebastian Jensen**

*Aarhus University and CREATEs*

### Abstract

We project carbon dioxide emissions through 2100, using a reduced-form model and national-level scenarios for per capita gross domestic product from the Shared Socioeconomic Pathways (SSPs). We propose a novel neural network-based panel data model that combines country fixed effects with a long short-term memory (LSTM) recurrent neural network regression component that takes into account time implicitly by building memory and letting model predictions depend on the income path of a country. For scenarios with low socioeconomic challenges for mitigation SSP1 and SSP4, our emissions projections appear consistent with baseline projections from structural integrated assessment models (IAMs) that are meant to describe future developments in absence of new climate policies. For scenarios with medium and high socioeconomic challenges for mitigation SSP2, SSP3, and SSP5, our emissions projections appear the most consistent with mitigation projections from IAMs that target a forcing level of 6.0 W/m<sup>2</sup> by 2100.

## 2.1 Introduction

Carbon dioxide (CO<sub>2</sub>) is the most important greenhouse gas and the key driver of climate change [IPCC] [2014]. Scenario-based CO<sub>2</sub> emissions projections help to determine whether current climate policies are sufficient for reaching policy goals such as those put forward by the Paris Agreement under the United Nations Framework Convention on Climate Change<sup>1</sup> and help to determine the expected costs of emissions reductions. It is necessary to project emissions several decades into the future, as most of today's CO<sub>2</sub> emissions remain in the atmosphere. It takes decades for the planet to respond to today's emissions especially due to thermal inertia of the deep oceans [IPCC] [2014]. Likewise, it may take decades to develop clean new technologies and to replace the capital and infrastructure associated with old technologies.

According to the Intergovernmental Panel on Climate Change of the United Nations, the increase in CO<sub>2</sub> emissions observed over the past decades has most likely been driven by income and population growth [IPCC] [2014]. This suggests that a model capturing the interrelationships between income, population, and emissions may be useful for projecting emissions into the future. These observations provide the impetus of the present paper, where we project CO<sub>2</sub> emissions through 2100 using national-level scenarios for per capita gross domestic product (GDP) from the Shared Socioeconomic Pathways (SSPs; Riahi et al. [2017]) and a reduced-form panel data model. Reduced-form models can be considered agnostic alternatives to the structural and computationally intensive integrated assessment models (IAMs) that are often used to make scenario-based emissions projections within climate change research. It seems important to also consider alternatives to IAMs as the computational burden associated with solving IAMs is heavy, they rely on a large number of structural and parametric assumptions, and parameter values are typically set using a mixture of judgment, a priori information, and calibration. By comparing our reduced-form emissions projections to those of popular IAMs, we assess whether the IAM projections appear consistent with historical experience. We want the employed model to be able to account for time-varying features of the relationship between per capita GDP and emissions. However, we want to avoid explicit dependence on time, as models that depend explicitly on time are difficult to use out-of-sample, as we discuss below. Therefore, we propose a novel neural network-based panel data model that can be readily used out-of-sample by avoiding explicit dependence on time, but which takes into account time implicitly by building memory and letting model predictions depend on the income path of a country. In particular, we extend the semiparametric panel data methodology developed by [Bennedsen, Hillebrand, and Jensen] [2021], and propose a model specification that combines country fixed effects with a long short-term memory (LSTM) recurrent neural network regression compo-

---

<sup>1</sup>The goal of the Paris Agreement is to limit the increase in global average temperature to well below 2 degrees Celsius above pre-industrial levels and to pursue efforts to limit the temperature increase to 1.5 degrees Celsius above pre-industrial levels [United Nations] [2015].

ment that uses per capita GDP as its only input variable, and is used to build memory. Alternative reduced-form panel data models from the environmental Kuznets curve (EKC) literature, initiated by Grossman and Krueger (1991), often take into account time explicitly through the use of time fixed effects, such as the quadratic model of Holtz-Eakin and Selden (1995), which has also been discussed by Wagner (2008 2015), the spline-based model of Schmalensee et al. (1998), and the static neural network model of Bennedsen et al. (2021). Models that rely on time fixed effects are difficult to use out-of-sample because no time fixed effects are available. Other EKC models depend explicitly on time through a nonparametric component, such as the functional coefficient approach of Chang, Choi, Kim, Miller, and Park (2016 2021) and the dynamic neural network model of Bennedsen et al. (2021). Such models are also difficult to use out of sample because there is no guarantee that the nonparametric model component has learned an appropriate relationship outside the support of the estimation sample. To motivate the use of our proposed model for constructing scenario-based emissions projections, we demonstrate in a pseudo out-of-sample experiment that our proposed model predicts both global and regional levels of emissions more accurately than the quadratic EKC model of Holtz-Eakin and Selden (1995) and the spline-based EKC model of Schmalensee et al. (1998), which have previously been considered for constructing scenario-based emissions projections.

We find that our CO<sub>2</sub> emission projections reflect socioeconomic challenges for mitigation. For scenarios with low socioeconomic challenges for mitigation SSP1 and SSP4, our projections appear consistent with baseline projections from IAMs that are meant to describe future developments in absence of new climate policies. For an intermediate scenario SSP2 and scenarios with high socioeconomic challenges for mitigation SSP3 and SSP5, our projections appear the most consistent with mitigation projections from IAMs that target a forcing level of 6.0 W/m<sup>2</sup> by 2100 through climate change mitigation policies. For all scenarios, our emissions projections are above IAM projections that target a forcing of 1.9 W/m<sup>2</sup> by 2100 which is broadly in line with reaching the goals of the Paris Agreement (Riahi et al. 2017).

The paper is structured as follows. In section 2.2 we discuss the literature on scenario-based CO<sub>2</sub> emissions projections and we also discuss econometric issues associated with reduced-form modeling of the income-emissions relationship. Section 2.3 discusses the data used for estimation and the scenarios used to project emissions into the future. Section 2.4 briefly presents the neural network-based panel data methodology developed in Bennedsen et al. (2021), then proposes a novel path-dependent neural network model specification. In Section 2.5 we compare out-of-sample predictive accuracy of the path-dependent neural network model to popular reduced-form alternatives from the EKC literature. In Section 2.6 we present scenario-based CO<sub>2</sub> emissions projections from the path-dependent neural network model and we assess whether projections made by popular IAMs appear consistent with historical experience. Section 2.7 briefly summarizes and concludes.

## 2.2 Scenario-based emissions projections and econometric issues

When constructing long-run emissions projections, the simplest approach is to extrapolate the historic growth rate; see Ausubel and Nordhaus (1983) for a review of this literature. This approach, however, does not take into account future responses to changes in the conditions of the economy, biosphere, or the atmosphere. A different approach, which explicitly takes into account interactions between economic processes and processes that produce greenhouse gases, is to use structural integrated assessment models<sup>2</sup> (IAMs) together with scenarios for future socioeconomic developments to assess future energy use and emissions; see Alcamo et al. (1995) for a review of this literature. However, the computational burden associated with solving these complex structural models is heavy and the models are not guaranteed to appropriately capture the real world. They rely on a large number of structural and parametric assumptions, and parameter values are typically set using a mixture of judgment and calibration. In this paper, we use an intermediate approach, where we estimate the reduced-form relationship between per capita gross domestic product (GDP) and carbon dioxide (CO<sub>2</sub>) emissions using a novel semiparametric panel data model that combines country fixed effects with a long short-term memory (LSTM) recurrent neural network regression component, then use scenarios from the Shared Socioeconomic Pathways (SSPs; Riahi et al. 2017) to construct future emissions projections. In contrast to extrapolate the historic growth rate, our approach embodies the historic interrelationships between economic development, population growth, and emissions. In contrast to structural IAMs, our approach is less computationally intensive, it avoids pre-imposing a lot of structural and parametric restrictions, and relies on parameter estimation that avoids use of judgment.

Holtz-Eakin and Selden (1995) were the first to use a reduced-form panel data model to construct long-run CO<sub>2</sub> emissions projections. They use national-level data to estimate the reduced-form relationship between per capita GDP and per capita CO<sub>2</sub> emissions using a quadratic emissions function augmented by country and time fixed effects. The paper projects emissions through 2100 using own GDP forecasts and population forecasts from *World population projections 1992-93* of the World Bank (World Bank 1992). Since time fixed effects are not available out-of-sample, Holtz-Eakin and Selden (1995) fix the value of the time fixed effects to the last value of the estimation sample in their main analysis, but explore sensitivity to this choice. They do not use scenarios for GDP and population per se, but they analyze sensitivity to the predicted rate of GDP growth. Schmalensee et al. (1998) replace the quadratic emissions function of Holtz-Eakin and Selden (1995) by linear splines and project CO<sub>2</sub> emissions though 2050 using the *IS92 scenarios* developed by the Intergovernmental Panel on Climate Change (IPCC) for the *IPCC Second Assessment Report* (IPCC 1995). Schmalensee et al. (1998) extrapolate their estimated time fixed effects into the

---

<sup>2</sup>Integrated assessment models are large-scale structural models that link key features of the economy with the key features of the biosphere and the atmosphere.

future under the assumption that climate policy will continue to be strengthened at roughly the historical pace, and find that emissions projections of the IPCC provide a significant departure from the implications of historical experience. Motivated by the observation that climate policy may have strengthened in pace since the adoption of the Kyoto protocol in 1997, Zhao and Du (2015) project CO<sub>2</sub> emission through 2050 using more recent scenarios and a more recent sample period, but consider only OECD and China. Zhao and Du (2015) use the *SRES scenarios* developed by the IPCC for the *IPCC Third Assessment Report* (IPCC 2001), and they use the quadratic model of Holtz-Eakin and Selden (1995), but extrapolate their estimated time fixed effects following the approach of Schmalensee et al. (1998). In this paper, we use the novel SSP scenarios to construct CO<sub>2</sub> emissions projections through 2100. The same scenarios are used by the IPCC in the *IPCC Sixth Assessment Report*<sup>3</sup> (IPCC 2021). An important feature of the SSPs is that population and GDP scenarios are available at a country level. Previous scenarios are available for aggregate macro-regions only. Availability of national-level data greatly facilitates panel data analysis.

The papers Holtz-Eakin and Selden (1995); Schmalensee et al. (1998); Zhao and Du (2015) (henceforth referred to as the reference papers) have evolved from the extensive environmental Kuznets curve (EKC) literature; see Bennedsen et al. (2021) for a literature review. The EKC refers to an inverse U-shaped relationship between income and emissions and dates back to the seminal contribution by Grossman and Krueger (1991). The EKC literature likely suffers from a variety of econometric issues. The most critical issues seem to be pertaining to functional misspecification, cross-sectional heterogeneity, structural changes, non-identifiable time effects, and integrated variables (Stern 2004; Müller-Fürstenberger and Wagner 2007; Wagner 2008). Below, we elaborate on these issues and how we seek to resolve them.

Pre-imposing a quadratic income-emissions function as in Holtz-Eakin and Selden (1995) and Zhao and Du (2015) may lead to functional misspecification (Harbaugh et al. 2002; Galeotti et al. 2006; Tsurumi and Managi 2015). Therefore, Schmalensee et al. (1998) drop the quadratic restriction and instead use linear splines for the income-emissions function. In our paper, we extend the neural network-based panel data methodology developed in Bennedsen et al. (2021) and propose to model the income-emissions function using an LSTM network. The use of an LSTM network has several advantages over the use of linear splines. We discuss some of them below; see also Bennedsen et al. (2021).

The reference papers assume the shape of the income-emissions relationship is common to all countries, and they allow only for intercept shifts across countries using country fixed effects. This assumption is likely overly restrictive (de Bruyn et al. 1998; List and Gallet 1999; Dijkgraaf and Vollebergh 2005; Bennedsen et al. 2021). In this paper, we use cross-country dependencies by having some model parameters be

---

<sup>3</sup>Since the *IPCC Fifth Assessment Report* (IPCC 2014), the IPCC has stepped down from developing its own scenarios. Instead, they rely on scenarios developed independently by the climate change research community.

common across all countries, and we allow for cross-country heterogeneity in the shape of the income-emissions relationship by having other model parameters vary across countries.

In the reference papers, estimation and model selection is based on a loss function defined over errors for a logarithmic transformation of per capita emissions. Auffhammer and Steinhauser (2012) argue the loss function should be defined over errors for the level of emissions when they are ultimately the object of interest. In this paper, we therefore use a squared error loss function defined over the level of emissions for both estimation and model selection.

In reduced-form models of the income-emissions relationship, separation of the effect of income from the effect of time hinges crucially on the identifying restrictions imposed by the researcher as both emissions and income are time dependent. The reference papers restrict the shape of the income-emissions relationship to be common to all time periods, and allow only for intercept shifts over time, using time fixed effects. Not only are time fixed effects not readily available out-of-sample, which complicates out-of-sample analysis, it may also be important to accommodate changes in the shape of the income-emissions relationship over time (Halkos and Tsionas 2001; Romero-Ávila 2008; Hendry 2018). Bennedsen et al. (2021) propose a dynamic neural network model to accommodate changes in shape over time by including a time variable as an additional input into a feedforward neural network, and Chang et al. (2016, 2021) propose the use of a panel data model with functional coefficients. However, models that depend explicitly on time through a nonparametric component are not guaranteed to be well-behaved for values of the time variable outside the estimation sample. In this paper, we propose a fundamentally different approach on how to utilize the time dimension of the data. We propose to use an LSTM recurrent neural network regression component that takes into account the time dimension of the data by allowing the shape of the income-emissions relationship to depend on the entire income path of a country. By avoiding explicit dependence on time, the model can be readily used out-of-sample. Recurrent neural networks denote a family of neural networks that are specialized in processing sequential data, such as time series data, by incorporating memory through feedback loops of information in the computational graph. The LSTM network was originally proposed by Hochreiter (1991); Gers, Schmidhuber, and Cummins (1999), and it has been much praised for its ability to learn the nonlinear and nonstationary nature of potentially integrated time series (Preeti, Bala, and Singh 2019). Faced with the challenge of learning long-term dependencies, recurrent neural networks often suffer from vanishing or exploding gradients, as gradients are propagated over many stages using backpropagation (Hochreiter 1991; Bengio, Frasconi, and Simard 1993; Bengio, Simard, and Frasconi 1994). The LSTM network solves these problems by introducing connection weights that may change at each time step (Goodfellow et al. 2016).

### 2.3 Data

Our data set contains national-level data on the following variables for the period 1960-2018:

- CO<sub>2</sub> emissions, megatonnes ( $10^6$  tonnes).
- GDP, billion 2005 U.S. dollars adjusted using purchasing power parities.
- Population, millions.

The data set contains 8,641 observations. It is an unbalanced panel of data with observations on 186 countries and balanced data for 81 countries. For the year 1960, the data set covers 86 countries, it accounts for 79.0% of the world's total population, and it accounts for 65.9% of the world's total CO<sub>2</sub> emissions. For the years 1990 onward, it covers more than 160 countries each year, and it accounts for more than 95.0% of the world's total population and CO<sub>2</sub> emissions.

Territorial CO<sub>2</sub> emissions estimates are from the Global Carbon Project (2019). Estimates include emissions from fossil fuel combustion, oxidation, and cement production, and exclude emissions from bunker fuels. The latter cannot be allocated unambiguously across countries. For the year 2018 (and for some countries 2016-2018), estimates are preliminary and made by the Global Carbon Project based on energy statistics published by British Petroleum. Data on population<sup>4</sup> and GDP are from the World Development Indicators database of the World Bank.<sup>5</sup> The GDP series we use is constructed from the separate series GDP<sup>6</sup> (current local currency units), the GDP deflator<sup>7</sup> and a purchasing power parity (PPP) conversion factor<sup>8</sup>.

In Table B.1 of the appendix, we map each country to one of five macro-regions, defined for the Shared Socioeconomic Pathways:

- OECD: 43 OECD90 and EU member states and candidates.
- REF: 13 reforming economies of Eastern Europe and the former Soviet Union.
- Asia: 35 Asian countries excluding the Middle East, Japan and former Soviet Union states.
- MAF: 64 countries of the Middle East and Africa.
- LAM: 34 countries of Latin America and the Caribbean.

---

<sup>4</sup>The series "SPPOPTOTL" was downloaded on December 12, 2019.

<sup>5</sup>Accessible at [databank.worldbank.org/source/world-development-indicators](https://databank.worldbank.org/source/world-development-indicators) last accessed on December 12, 2019.

<sup>6</sup>The series "NY.GDP.MKTP.CN" was downloaded on December 12, 2019.

<sup>7</sup>The series "NY.GDP.DEFL.ZS" was downloaded on December 12, 2019.

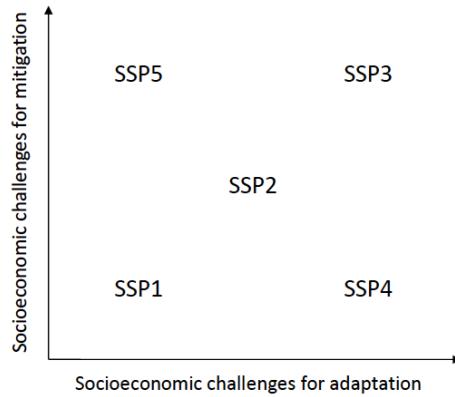
<sup>8</sup>The series "PA.NUS.PPP" was downloaded on December 12, 2021.

Tables B.2 and B.3 of the appendix report descriptive statistics for the regions. Table B.2 reports the mean of per capita GDP and per capita emissions. We note that OECD is the richest region throughout the sample period. Asia and MAF are the poorest regions. Table B.3 reports the standard deviation of per capita GDP and per capita emissions. The standard deviation of per capita GDP and per capita emissions are the largest for the regions OECD and MAF which are also composed of the largest number of individual countries. Hence, it seems these regions may be the most heterogeneous regions.

### 2.3.1 Shared Socioeconomic Pathways

We use scenarios from the Shared Socioeconomic Pathways (SSPs) to construct future emissions projections. The same scenarios are used by the IPCC for the IPCC Sixth Assessment Report<sup>9</sup> [IPCC 2021]. The SSPs have been developed over the last years through a joint community effort, and provide distinct scenarios for socioeconomic developments. The framework of the Shared Socioeconomic Pathways consists of five main scenarios SSP1-SSP5 that contain a narrative, national-level GDP, population, and urbanization trajectories, and qualitative assumptions on the energy and land-use sectors (Riahi et al. 2017). The five main scenarios span the space of socioeconomic challenges to mitigation and adaptation, as illustrated in Figure 2.1. Socioeconomic challenges to mitigation consist of factors that lead to higher reference emissions in absence of new climate policy, and factors that tend to reduce mitigative capacity of society. According to O'Neill et al. (2014), factors that could contribute to higher reference emissions in the SSPs include high population growth rates, rapid economic growth, extensive land use, energy intensive economic systems, and carbon intensive energy supplies; and factors that tend to influence mitigative capacity of society include the range of viable technological options, policy institutions, availability of financial resources necessary to support mitigative activities, and the stock of human and social capital. Socioeconomic challenges to adaptation consist of factors that increase risks associated with climate change. According to O'Neill et al. (2014), socioeconomic challenges to adaptation are a function of socioeconomic determinants of exposure to climate change hazards, such as sea level rise, changes in temperature and precipitation, extreme events, sensitivity of these hazards, and the capacity of society to cope with these hazards. We expect our reduced-form model, presented in the next section, best capture challenges to mitigation, as most factors affecting challenges to adaptation are left unmodeled. We show global trajectories for GDP, population, per capita GDP, and inequality across countries in Figure 2.2. We measure inequality across countries using the ratio between median income and the upper bound value of the first decile. This is one of the income inequality ratios considered by the Organization for Economic Cooperation and Development

<sup>9</sup>The IPCC uses a different variant of the scenarios from climate models participating in the Coupled Model Intercomparison Project Phase 6 (CMIP6) of the World Climate Research Programme [IPCC 2021].

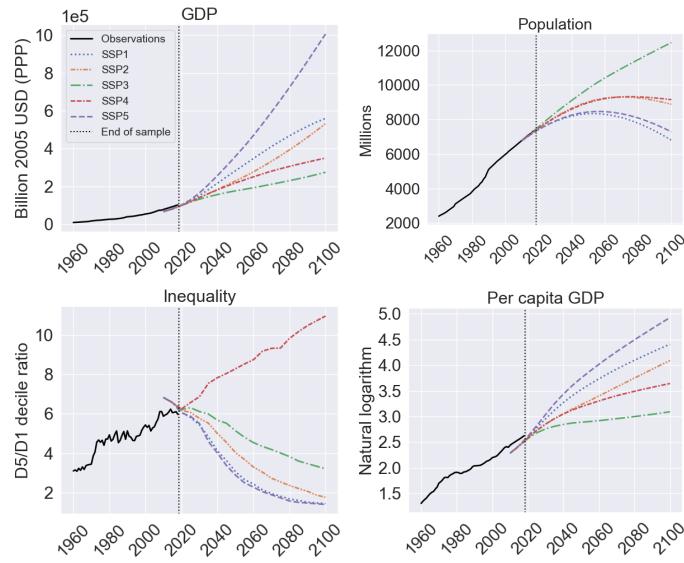
**Figure 2.1:** The Shared Socioeconomic Pathways

Source: Adapted from Dellink et al. (2017).

(OECD; OECD 2018). The paper Dellink, Chateau, Lanzi, and Magné (2017) instead uses the top-bottom decile income ratio to compare inequality across the SSPs. We show trajectories for each region separately in Figure B.1 of the appendix. The plots in Figure B.1 are qualitatively similar to those in Figure 2.2. Summaries of the SSP narratives are provided in Riahi et al. (2017).

SSP1 describes a world characterized by sustainability and reduced fossil fuel dependency (Riahi et al. 2017). SSP1 is also characterized by low socioeconomic challenges for both mitigation and adaptation (Figure 2.1) and it has the second highest income trajectories and low inequality (Figure 2.2). SSP4 is characterized by low socioeconomic challenges for mitigation, but high socioeconomic challenges for adaptation (Figure 2.1). SSP4 also has the largest degree of inequality between countries (Figure 2.2). SSP3 describes a world characterized by regional rivalry and little progress in reducing fossil fuel dependency (Riahi et al. 2017). SSP3 is characterized by high socioeconomic challenges for both mitigation and adaptation (Figure 2.1); it is the only scenario with a continuously increasing global population and it has the lowest income trajectories (Figure 2.2). Also SSP5 describes a world dominated by fossil fuels, leading to high challenges for mitigation (Riahi et al. 2017). In addition, we note that SSP5 is characterized by high-income trajectories and low inequality (Figure 2.2). Finally, SSP2 is a “middle of the road” scenario characterized by intermediate trajectories for income, population, and inequality (Figure 2.2). It is meant to describe a world in which social, economic, and technological trends do not shift markedly from historical trends (Riahi et al. 2017).

One might worry that the SSP scenarios look so far ahead in time that the estimation sample from 1960-2018 contains little useful information. That is not necessarily the case. For each region (including global) and scenario, Table 2.1 reports the frac-

**Figure 2.2:** Shared Socioeconomic Pathways basic elements: global level

Note: In the bottom left panel, the expression D5/D1 is the ratio of median income to the upper bound value of the first decile. Income is measured by per capita GDP.

tion of national-level values of per capita GDP that are within the minimum and maximum value observed for any given country within that region in the estimation sample. Note that the model we propose (in the next section) assumes the shape of the relationship between per capita GDP and emissions is common to countries within a given region. It is therefore of little concern that the value of per capita GDP for a given country is outside the range of values observed for that country in the estimation sample if we have observed values of per capita GDP at least as high and at least as low for at least one other country within that region in the estimation sample; in which case we say that observation is within sample range for that region. For the scenarios SSP1-SSP4, essentially all observations are within global sample range through 2100. For SSP5, 87% of the observations are within global sample range by 2100, and the percentage remains close to 100 until near 2100. As expected, the percentage of observations within sample range is decreasing over time. We also note from Table 2.1 that the percentage of observations within sample range varies across regions. However, as the percentage of observations within global sample range remains close to 100 through 2100, most observations out of sample range for some regions must be within sample range for other regions. Therefore, it is likely also important that the model we use to learn the shape of the relationship between per capita GDP and emissions is able to use information from across regions. The model we propose uses information from across regions by having some model parameters

**Table 2.1:** Per capita income in the Shared Socioeconomic Pathways within sample range

	2020	2040	2060	2080	2100
Global					
SSP1	1.00	1.00	1.00	1.00	1.00
SSP2	1.00	1.00	1.00	1.00	1.00
SSP3	1.00	1.00	1.00	1.00	1.00
SSP4	1.00	1.00	1.00	0.99	0.99
SSP5	1.00	0.99	0.99	0.99	0.87
OECD					
SSP1	1.0	0.98	0.93	0.46	0.17
SSP2	1.0	0.98	0.89	0.54	0.22
SSP3	1.0	0.98	0.98	0.88	0.66
SSP4	1.0	0.98	0.85	0.39	0.20
SSP5	1.0	0.98	0.61	0.15	0.00
REF					
SSP1	1.00	0.50	0.17	0.00	0.00
SSP2	1.00	0.67	0.42	0.00	0.00
SSP3	1.00	0.67	0.67	0.42	0.33
SSP4	1.00	0.67	0.25	0.17	0.17
SSP5	1.00	0.50	0.00	0.00	0.00
Asia					
SSP1	1.00	1.00	1.00	1.00	1.00
SSP2	1.00	1.00	1.00	1.00	0.97
SSP3	1.00	1.00	1.00	1.00	1.00
SSP4	1.00	1.00	1.00	1.00	0.97
SSP5	1.00	1.00	1.00	0.87	0.47
MAF					
SSP1	1.00	1.00	1.00	1.00	1.00
SSP2	1.00	1.00	1.00	1.00	1.00
SSP3	1.00	1.00	1.00	1.00	1.00
SSP4	1.00	1.00	1.00	0.98	0.98
SSP5	1.00	0.98	0.98	0.97	0.90
LAM					
SSP1	1.00	1.00	0.76	0.38	0.00
SSP2	1.00	1.00	0.97	0.72	0.21
SSP3	1.00	1.00	1.00	1.00	1.00
SSP4	1.00	1.00	0.97	0.69	0.38
SSP5	1.00	1.00	0.55	0.00	0.00

Note: Numbers reflect fractions of values of per capita GDP in the Shared Socioeconomic Pathways (SSPs) that are within the minimum and maximum value observed in the estimation sample.

**Table 2.2:** SSP scenario matrix framework

Forcing (W/m <sup>2</sup> )	SSP1	SSP2	SSP3	SSP4	SSP5
Baseline	SSP1-Base	SSP2-Base	SSP3-Base	SSP4-Base	SSP5-Base
6	NA	SSP2-60	SSP3-60	SSP4-60	SSP5-60
4.5	SSP1-45	SSP2-45	SSP3-45	SSP4-45	SSP5-45
3.4	SSP1-34	SSP2-34	SSP3-34	SSP4-34	SSP5-34
2.6	SSP1-26	SSP2-26	NA	SSP4-26	SSP5-26
1.9	SSP1-19	SSP2-19	NA	NA	SSP5-19

Source: Adapted from Riahi et al. [2017].

be common to all countries while other parameters are specific to countries within a given region.

From the five main scenarios SSP1-SSP5, different baseline and mitigation scenarios are derived. The baseline scenarios serve as reference scenarios and are meant to describe future developments in absence of new climate policies beyond those already in place. The mitigation scenarios consider implications of climate change mitigation policies (Riahi et al. [2017]). This leads to a scenario matrix framework illustrated in Table 2.2. The full set of baseline and mitigation scenarios are derived by elaborating the five main scenarios using IAMs (Fricko et al. [2017] Fujimori et al. [2017] Calvin et al. [2017] Kriegler et al. [2017]). The mitigation scenarios are derived by coupling the main scenarios with target forcing levels<sup>10</sup> for 2100 from the representative concentration pathways (RCPs; van Vuuren et al. [2011]). Three of the four target forcings from the RCPs<sup>11</sup> are used if applicable: 6.0, 4.5, 2.6 W/m<sup>2</sup>. The target forcings 6.0 and 4.5 W/m<sup>2</sup> reflect medium stabilization scenarios with intermediate efforts and policies to reduce emissions. The target forcing 2.6 W/m<sup>2</sup> requires stringent efforts to reduce emissions and is broadly consistent with the objective of limiting global average temperature to 2 degrees Celsius above pre-industrial levels. The RCP target forcing 8.5 W/m<sup>2</sup> is not used explicitly in the SSP scenario matrix framework. However, one of the SSP5 baseline scenarios reaches a forcing level for 2100 around 8.5 W/m<sup>2</sup><sup>12</sup>. In addition, the SSP scenario matrix framework considers an intermediate target forcing of 3.4 W/m<sup>2</sup> and a very low target forcing of 1.9 W/m<sup>2</sup>. The target forcing 1.9 W/m<sup>2</sup> is motivated by the Paris Agreement and its attention to the impor-

<sup>10</sup> Anthropogenic radiative forcing measures the net change in the energy balance of the Earth system due to human activities relative to pre-industrial levels, typically expressed in watts per square meter (W/m<sup>2</sup>; Myhre et al. [2013]). It is a measure of human contribution to climate change.

<sup>11</sup> The exact target forcings used in the SSP scenario matrix framework of some RCPs differ slightly from the original RCP labels. Based on the climate model MAGICC-6, the target forcings for 2100 used in the SSP scenario matrix framework are 5.4, 4.2, 2.6 W/m<sup>2</sup> (Riahi et al. [2017]).

<sup>12</sup> This particular SSP5 baseline scenario is used in the CMIP6 under the label "SSP5-85" in replacement of RCP8.5 used in the CMIP5.

tance of exploring temperature levels even lower than 2 degrees Celsius [Riahi et al. 2017]. Each of the baseline and mitigation scenarios are implemented using multiple IAMs. In total, six IAMs are used: IMAGE [Bouwman, Kram, and Klein Goldewijk 2006], MESSAGE-GLOBIOM [Messner and Strubegger 1995], AIM/CGE [Fujimori, Masui, and Yuzuru 2012], GCAM4 [Calvin 2011], REMIND-MAGPIE [Luderer et al. 2013], and WITCH-GLOBIOM [Emmerling et al. 2016]. For each SSP, a so-called *marker* scenario is selected as representative of the broader developments of that SSP (i.e. a marker model is selected for each of the columns in Table 2.2). For SSP1-SSP5, the marker scenarios are from IMAGE, MESSAGE-GLOBIOM, AIM/CGE, GCAM4, and REMIND-MAGPIE, respectively. As remarked by [Riahi et al. 2017], the marker scenarios do not provide a median estimate, and *non-marker* scenarios are important, as they provide a first-order estimate of conditional uncertainties pertaining to model structure and interpretation of the SSPs. Not all scenarios are reachable by every IAM, and few scenarios are reachable by no IAMs (see Table 2.2). Finally, greenhouse gas and aerosol emissions from the IAMs are translated into concentrations, forcings, and other climate change-related variables using the MAGICC-6 simple carbon cycle climate model [Riahi et al. 2017].

## 2.4 Methodology

In this section, we briefly introduce the methodology developed in [Bennedsen et al. 2021], then present a novel path-dependent neural network model specification. This section therefore follows the exposition in [Bennedsen et al. 2021] closely. The reduced-form relationship between income and greenhouse gas emissions can be mathematically represented in a panel data framework as

$$y_{it} = f(x_{it}, i, t) + u_{it}, \quad i = 1, \dots, N_t; t = 1, \dots, T, \quad (2.1)$$

where  $y$  is a measure of emissions (here  $\text{CO}_2$ ),  $x$  is an income measure, and  $u$  is an error term. Here,  $i$  indexes countries and  $t$  indexes time periods. Equation (2.1) allows for an unbalanced panel of data by letting  $N_t \leq N$  denote the number of countries observed in time period  $t$  of the total number of countries  $N$ . In most studies of the income-emissions relationship,  $y$  is log-transformed per capita emissions and  $x$  is log-transformed per capita GDP. The focus on per capita quantities reflects the hypothesis that population sizes do not affect average behavior. The log-linear specification is typically preferred over a linear specification as multiplicative country and time specific effects are deemed more plausible than additive effects given the heterogeneity of countries in a typical study [Schmalensee et al. 1998]. The function  $f$  characterizes the functional relationship between income and emissions. Without imposing further restrictions,  $f$  cannot be identified from equation (2.1), as only one observation pair  $(y_{it}, x_{it})$  is available for each combination of country and time period  $(i, t)$ . Unfortunately, it is not clear which identifying restrictions

are appropriate. As noted by Vollebergh et al. (2009), a particularly important issue relates to the fact that income and emissions are both time dependent. Separation of the income and the time effect is therefore dependent on the identifying restrictions imposed by the researcher. If one allows fully flexible time effects that are also country-specific, all variation in the data will be explained by these control variables. If one overly restricts the time and country effects, too much of the explained variation will be attributed to the income effect.

In the literature on the environmental Kuznets curve (Holtz-Eakin and Selden 1995; Wagner 2008, 2015), it is standard to identify the function  $f$  in equation (2.1) by imposing the restriction that it is quadratic and common to all countries and time periods,  $f(x_{it}, i, t) = f(x_{it})$ , and that  $\nu$  is a composite error term consisting of additively separable county and time fixed effects in addition to an idiosyncratic and purely stochastic effect. The quadratic EKC model can be represented as

$$y_{it} = \alpha_i + \beta_t + \delta_1 x_{it} + \delta_2 x_{it}^2 + \nu_{it}, \quad (2.2)$$

where  $\alpha$  and  $\beta$  are typically treated as fixed effects, and  $\nu$  is a stochastic error term. The fixed effects are supposed to control for omitted variables that are correlated with income, but which are not endogenous consequences of income changes. The country fixed effects  $\alpha$  are supposed to capture exogenous and persistent cross-country differences in features such as fossil fuel availabilities and prices, output mixes, regulatory structures, policies, and tastes. The time fixed effects  $\beta$  are supposed to capture time-varying omitted variables and shocks which are common to all countries (Stern 2017). They capture the effect on emissions over time in absence of changes in income and can likely be interpreted as capturing the effect of common technology shocks that is not captured by income changes (Bennedsen et al. 2021).

Although equation (2.2) has been extensively applied within the EKC literature, the appropriateness of it has been debated (Stern 2004, 2017; Müller-Fürstenberger and Wagner 2007; Wagner 2008). Bennedsen et al. (2021) propose a number of novel identifying restrictions for  $f$  in (2.1), and consider the implications for the shape of the estimated income-emissions relationship. Most importantly, Bennedsen et al. (2021) drop the parametric restriction that  $f$  is necessarily quadratic and common to all countries, and represent  $f$  nonparametrically using a feedforward neural network. The paper proposes two distinct model specifications:

$$y_{it} = \alpha_i + \beta_t(r) + f^{\text{FNN}}(x_{it}, r) + \nu_{it}, \quad (\text{Static neural network model})$$

$$y_{it} = \alpha_i + f^{\text{FNN}}(x_{it}, t, r) + \nu_{it}, \quad (\text{Dynamic neural network model})$$

where the superscript on  $f$  highlights the fact that a feedforward neural network is used to represent  $f$  rather than pre-imposing a quadratic relationship. Here,  $r$  is a regional indicator. A function  $\tau : \{1, \dots, N\} \rightarrow \{1, \dots, R\}; i \mapsto r$ ,  $1 \leq R \leq N$ , is used to map each country to a region of homogeneous countries within which countries are

assumed to admit the same functional relationship (up to country-specific intercept shifts) and potentially time effects. [Bennedsen et al. (2021)] document good in-sample performance of these model. However, as we discuss below, these models are not necessarily well suited for out-of-sample analysis.

The static model specification cannot readily be used for projecting emissions into the future, as no time fixed effects are available for periods beyond the estimation sample. The quadratic EKC model in (2.2) and other semiparametric alternatives such as the spline-based EKC model of [Schmalensee et al. (1998)] face similar issues. One approach to solving this issue is to use the estimated time fixed effect for the last period of the estimation sample. This is the approach suggested by [Holtz-Eakin and Selden (1995)]. The paper by [Schmalensee et al. (1998)] suggests to instead project time fixed effects into the future by fitting a simple parametric function (such quadratic or piecewise linear) to the time fixed effects for the estimation sample, then use that function to extrapolate out-of-sample. Using the approach of [Schmalensee et al. (1998)], time fixed effects quickly start dominating the emissions projections. The choice of function used to project time fixed effects into the future is therefore not innocent, and as discussed by [Holtz-Eakin and Selden (1995)], it is often not clear what is an appropriate function. Thus, we believe a model with explicit time fixed effects is not well suited for projecting emissions into the future.

The dynamic neural network model specification of [Bennedsen et al. (2021)] does not require projecting separable time fixed effects into the future in an initial step. If the estimated relationship were to be reflective of the true functional relationship of interest over most of the input space, the model would be appropriate for making emissions projections out-of-sample. However, neural network models are so flexible that they typically learn a close approximation to the functional relationship of interest only in the region of input space where the model has data available for estimation [Bennedsen et al. (2021)]. When projecting emissions into the future, one would inevitably have to evaluate the model in values of the time variable  $t$  from outside the input region used for estimation. For this reason, we do not think the dynamic neural network model is well suited for projecting emissions into the future. The functional coefficient approach of [Chang et al. (2016) (2021)] face similar issues.

In this paper, we propose a novel model specification that can be readily used for projecting emissions into the future. The key distinguishing feature of this model is that it avoids explicit dependence on time. Instead, it takes into account time implicitly through a memory component by letting model predictions depend on the income path of a country. The memory component of the model is *country-specific* in the sense that output for a given country is based on the income path for that country only. We propose a path-dependent neural network model that represents  $f$  in (2.1) nonparametrically using a long short-term memory (LSTM) recurrent neural

network:

$$y_{it} = \alpha_i + f^{\text{LSTM}}(x_{it}, x_{it-1}, \dots, x_{i1}, r) + v_{it},$$

(Path-dependent neural network model)

where  $r$  is again a regional indicator. The superscript on  $f$  highlights the fact that we use an LSTM network to represent  $f$ . Here,  $x_{it}, x_{it-1}, \dots, x_{i1}$  is used to indicate path dependence. It does not mean the model has separate parameters for each element in the income sequence in the same way as a high-order autoregressive model or a feedforward neural network working on a sequence of fixed length. Rather, the term  $f^{\text{LSTM}}(x_{it}, x_{it-1}, \dots, x_{i1}, r)$  is meant to reflect the *unfolded* computational graph from a complex recurrence relation (by recursive substitution into itself) in variables derived from  $x$ , as we discuss below. The model is therefore non-Markovian in  $x$  itself. For a given value of current income, the model prediction of emissions may therefore be different for a country that is, say, on an increasing income path, and a country that is, say, on a decreasing income path. Because the model is specified by a recurrence relation, it has the same input size regardless of the length of the input sequence. It can therefore be used across countries and for sequence lengths not observed in the estimation sample. This is convenient when used to make out-of-sample predictions in Section 2.5 and scenario projections in Section 2.6 where the input sequence used for prediction differs from that used for estimation.

### 2.4.1 Path-dependent neural network model

This section discusses the proposed path-dependent neural network model in detail. The model imposes no parametric restrictions on the functional form of  $f$  in [2.1]. Instead, it learns  $f$  directly from data using a long short-term memory (LSTM) recurrent neural network. It assumes the error term  $u$  in [2.1] is a composite error term consisting of additively separable country fixed effects, in addition to an idiosyncratic and purely stochastic effect. By using an LSTM network architecture for the nonparametric component of the model, it allows the shape of the functional relationship between income and emissions to depend on the path of income values of a country, as we discuss below. Some parameters of the model are common to all countries and other parameters are allowed to vary across pre-specified regions. We initially map each country to a region of homogeneous countries according to  $\tau : \{1, \dots, N\} \rightarrow \{1, \dots, R\}; i \mapsto r, 1 \leq R \leq N$ . The model can then be mathematically

represented as

$$y_{it} = \alpha_i + \phi(r)^\top z_{it}^{(h)} + \nu_{it}, \quad i = 1, \dots, N_t; t = 1, \dots, T; r = \mathbf{r}(i), \quad (2.3)$$

$$z_{it}^{(h)} = \mathcal{H}^{(h)}(z_{it}^{(h-1)}, z_{it-1}^{(h)}), \quad (2.4)$$

⋮

$$z_{it}^{(2)} = \mathcal{H}^{(2)}(z_{it}^{(1)}, z_{it-1}^{(2)}), \quad (2.5)$$

$$z_{it}^{(1)} = \mathcal{H}^{(1)}(x_{it}, z_{it-1}^{(1)}), \quad (2.6)$$

where  $x_{it}, y_{it} \in \mathbb{R}$  are the natural logarithm of, respectively, per capita GDP and per capita emissions of country  $i$  in year  $t$ . The vector  $z_{it}^{(\ell)} \in (-1, 1)^{q_\ell}$ ,  $\ell = 1, \dots, h$ , is a column vector of derived variables by the LSTM network component of the model, and  $\phi(r) \in \mathbb{R}^{q_h}$  is a column vector of region-specific slope coefficients to be estimated from data. As we discuss below, it is common practice to *squash* the derived variables to the interval  $(-1, 1)$ , but one could allow the derived variables to take on values on the entire real line. Note that (2.3), which we will refer to as the output layer of the model, is a linear regression model in the derived variables  $z_{it}^{(h)}$ , augmented by country fixed effects  $\alpha_i$ . The derived variables constitute nonlinear transformations of current and past input variables,  $x_{it}$ , learned through a sequence of LSTM layers given by (2.4)-(2.6), referred to as *hidden layers*. Through the use of LSTM layers, the model is non-Markovian in  $x_{it}$ , and this is the sense in which the model is *path-dependent*. As we discuss below, the nonlinear transformations that are applied to  $x_{it}$  are assumed to be common across all countries, and we use the entire panel of data to learn how to optimally transform the input variable  $x_{it}$  into  $z_{it}^{(h)}$ . However, note that the slope coefficients of the output layer  $\phi(r)$  are region-specific,  $1 \leq r \leq R$ , to allow for emissions predictions and the effect of being on a given income path to differ across regions. Different choices of  $R$  provide different bias-variance tradeoffs by relying on different degrees of parameter sharing across countries. The paper by Bennedsen et al. (2021) detects important differences in the shape of the income-emissions relationship across the five regions discussed in Section 2.3: OECD, REF, Asia, MAF, and LAM. In the experiments of Section 2.5 and Section 2.6 we, therefore, consider  $R = 1$  and  $R = 5$ , where the five regions considered are those of Section 2.3. One of the key features of the LSTM layers is that they learn to decide for themselves the amount of current and past information that should be used to construct the derived  $z_{it}^{(h)}$ , using a sequence of *gates*, which we discuss below. In equations (2.3) and (2.4),  $h \geq 1$  denotes the number of layers used to learn the derived variables and is said to determine the *depth* of the network;  $q_\ell$  denotes the dimensionality of the  $\ell$ -th layer and is referred to as *width*. The choice of model depth and widths constitutes important architectural considerations with implications for the representation capabilities of the model and estimation difficulty. Much like the universal approximation capabilities of a feedforward neural network with a single hidden layer (Hornik et al. 1989, Cybenko 1989, Leshno et al. 1993), essentially any

function involving recurrence can be modeled using a recurrent neural network like the LSTM network, using a single hidden layer of finite size [Siegelmann and Sontag 1991] [1995] [Siegelmann 1995] [Hyötyniemi 1996]. However, multiple hidden layers may benefit in practice [Graves, rahman Mohamed, and Hinton 2013]. In Section 2.4.3 we discuss how we decide on an optimal network architecture. The function  $\mathcal{H}^{(\ell)}$  is used to implement the  $\ell$ -th hidden layer. As mentioned, we use the LSTM network for all hidden layer functions and implement  $\mathcal{H}^{(\ell)}$ ,  $\ell = 1, \dots, h$ , using the following composite function:

$$z_{it}^{(\ell)} = o_{it}^{(\ell)} \odot \tanh(s_{it}^{(\ell)}), \quad (2.7)$$

$$s_{it}^{(\ell)} = f_{it}^{(\ell)} \odot s_{it-1}^{(\ell)} + g_{it}^{(\ell)} \odot \tanh\left(\kappa_s^{(\ell)} + \Gamma_{z^{(\ell-1)}s}^{(\ell)} z_{it}^{(\ell-1)} + \Gamma_{z^{(\ell)}s}^{(\ell)} z_{it-1}^{(\ell)}\right), \quad (2.8)$$

$$f_{it}^{(\ell)} = \sigma\left(\kappa_f^{(\ell)} + \Gamma_{z^{(\ell-1)}f}^{(\ell)} z_{it}^{(\ell-1)} + \Gamma_{z^{(\ell)}f}^{(\ell)} z_{it-1}^{(\ell)}\right), \quad (2.9)$$

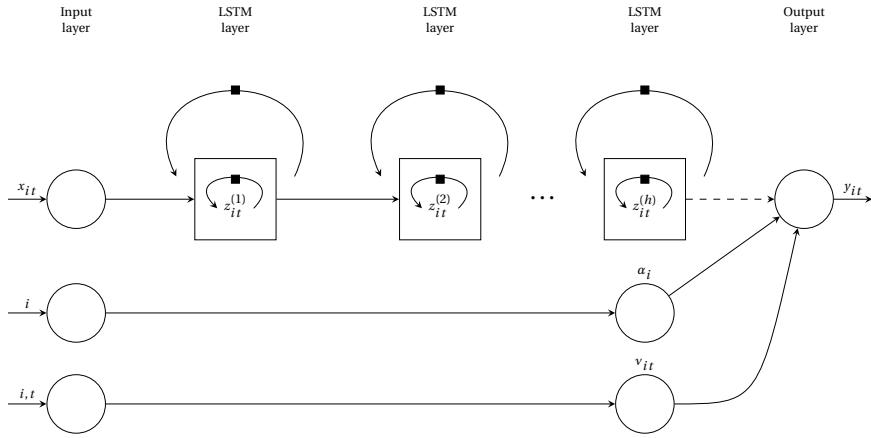
$$g_{it}^{(\ell)} = \sigma\left(\kappa_g^{(\ell)} + \Gamma_{z^{(\ell-1)}g}^{(\ell)} z_{it}^{(\ell-1)} + \Gamma_{z^{(\ell)}g}^{(\ell)} z_{it-1}^{(\ell)}\right), \quad (2.10)$$

$$o_{it}^{(\ell)} = \sigma\left(\kappa_o^{(\ell)} + \Gamma_{z^{(\ell-1)}o}^{(\ell)} z_{it}^{(\ell-1)} + \Gamma_{z^{(\ell)}o}^{(\ell)} z_{it-1}^{(\ell)}\right), \quad (2.11)$$

where  $z_{it}^{(\ell)}, s_{it}^{(\ell)}, \in (-1, 1)^{q_\ell}$ ,  $f_{it}^{(\ell)}, g_{it}^{(\ell)}, o_{it}^{(\ell)} \in (0, 1)^{q_\ell}$  are column vectors,  $\ell = 1, \dots, h$ , and the operator  $\odot$  is the Hadamard product. In equations (2.8)-(2.11), we follow the convention to set  $z_{it}^{(0)} \equiv x_{it}$ . We can think of the LSTM layers as consisting of so-called *memory cells*. The most important component of a memory cell is its *cell state*  $s$ , defined in (2.8). It is through the cell states that emissions predictions of the model depend on the income path of a country. The cell state allows a memory cell to integrate information from the past and remembers input values over arbitrary time intervals<sup>13</sup>. Some memory cells may work at fine-grained time scales and handle small details. Others may work at coarse time scales and transfer information from the past to the present more efficiently. Note from (2.8) that the cell state has an internal recurrence in addition to the outer recurrence of the model that feeds the output of the LSTM layer into itself with a one-period time lag. The weight applied to the internal recurrence is controlled by so-called *forget gate*  $f$ , defined in (2.9). The forget gate removes information from the cell state that is no longer deemed useful. It is context dependent, which allows the time scale of integration to be changed dynamically based on the input sequence. The flow of information into the cell state is controlled by a so-called external *input gate*  $g$ , defined in (2.10). In equation (2.8), the cell state is presented with current output from the preceding layer and its own lagged value. However, only some of that information is allowed by the external input gate to enter into the current cell state. The flow of information out of the cell state is controlled by a so-called *output gate*  $o$ , defined in (2.11). In (2.7), we note the derived

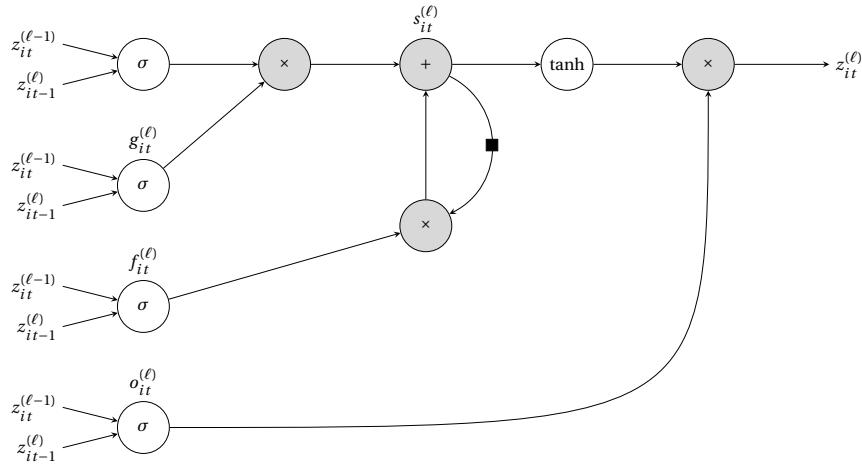
---

<sup>13</sup>To understand the name “long short-term memory network”, note that activations of a neural network can be understood to reflect short-term memory and parameters can be understood to reflect long-term memory. In an LSTM network, activations can preserve information over long time spans, which in a sense makes them *long short-term memory*.

**Figure 2.3:** Path-dependent neural network model

Note: Black squares indicate a lag of one time period. The dashed edge in the upper part of the graph represent region-specific connections.

variables  $z_{it}^{(\ell)}$  are constructed by applying the output gate to the current cell state. We use the hyperbolic tangent function in (2.7) and (2.8), as it seems the most popular choice in the literature (Graves et al., 2013) (Goodfellow et al., 2016). In principle, any so-called *activation function* could be used. Other popular choices for the activation

**Figure 2.4:** LSTM layer  $\ell$ 

Note: The black square indicates a lag of one time period.

function include the rectified linear unit<sup>14</sup> (ReLU; Glorot et al. 2011) activation function and the Swish activation function<sup>15</sup> (Ramachandran et al. 2017) in which case the elements of the derived variables and the cell states would not be constrained to the interval  $(-1, 1)$ . The activation function  $\sigma$  used in (2.9)-(2.11) is the logistic sigmoid,  $\sigma(z) = (1 + \exp(-z))^{-1}$  for  $z \in \mathbb{R}$ , implying that all gating values are between 0 and 1. Note that all activation functions are applied elementwise in equations (2.7)-(2.11). The vector  $\kappa_j^{(\ell)} \in \mathbb{R}^{q_\ell}$  is a column vector of intercepts<sup>16</sup> to be estimated from data; and the matrices  $\Gamma_{z^{(\ell)} j}^{(\ell)} \in \mathbb{R}^{q_\ell \times q_\ell}$  and  $\Gamma_{z^{(\ell-1)} j}^{(\ell)} \in \mathbb{R}^{q_\ell \times q_{\ell-1}}$  contain slope coefficients to be estimated from data,  $\ell = 1, \dots, h$ ,  $j = s, f, g, o$ . Note that the parameters of the LSTM layers are assumed to be common to all countries. In this way, we exploit cross-country dependencies to learn how to optimally transform the input variable  $x_{it}$  into  $z_{it}^{(h)}$  by using the entire panel of data.

The computational graph associated with the path-dependent neural network model is visually represented in Figure 2.3. The upper part of the figure illustrates the LSTM network component of the model. The bottom part of the graph illustrates the parametric component of the model. Edges represent how information travels through the model. The model is said to be recurrent, because the LSTM component includes feedback loops of information (recurrences). Vertices of the input layer reflect information that is presented to the model. In the LSTM component of the model, per capita income is passed through a sequence of LSTM layers. Figure 2.4 illustrates the computational graph associated with an LSTM layer. Each LSTM layer includes inner recurrences within cell states in addition to an outer recurrence. The inner recurrences are used to accumulate information and build memory in accordance with the discussion above.

### 2.4.2 Estimation

The intended use of the path-dependent neural network model is to predict global levels of CO<sub>2</sub> emissions (not the log-level of per capita emissions which is the dependent variable). Global levels of emissions is what affects concentrations in the atmosphere, and thereby temperature changes. As discussed by Auffhammer and Steinausser (2012), the loss function used for estimation should reflect the intended use of the model. Therefore, we estimate all free parameters of the model simultaneously by minimizing a global mean squared error loss function, where errors are based on the distance between observed and predicted levels of emissions:

$$J(\phi, \kappa, \Gamma, \alpha) = \sum_{r=1}^R \sum_{t=1}^T \sum_{i \in I_r} \frac{1}{n} \left( [e^{y_{it}} - e^{\hat{y}_{it}(x_{it})}] \times \text{POP}_{it} \right)^2, \quad (2.12)$$

---

<sup>14</sup>The ReLU function is defined as  $g(z) = \max(z, 0)$  for  $z \in \mathbb{R}$ .

<sup>15</sup>The Swish function is defined as  $g(z) = z(1 + \exp(-z))^{-1}$  for  $z \in \mathbb{R}$ .

<sup>16</sup>Typically referred to as *biases* in the neural network literature.

where  $I_r \subseteq \{1, 2, \dots, N\}$  is the set of indices of countries belonging to region  $r$ ,  $T$  is the number of sample periods,  $n$  is the total number of observations across countries, and POP is the population size. Note the model output  $\hat{y}$  is dependent on the estimated parameter vectors  $\hat{\phi}$ ,  $\hat{\kappa}$  and  $\hat{\Gamma}$ ,  $\hat{\alpha}$ , which are understood to contain parameters of all layers and regions.

Before discussing the optimization routine used to minimize the loss function in (2.12), we elaborate on a subtlety regarding identification of the country fixed effects. As discussed by Bennedsen et al. (2021), neural network models are so flexible that they automatically adapt their predictions to the level of the dependent variable in the estimation sample. From an illustrative point of view, neural network models can be thought of as implicitly identifying an overall intercept even if not specified explicitly. This is true also if imposing restrictions such as  $f(0) = 0$ . Therefore, we can only identify country fixed effects for all but one country. We use the U.S. as reference when estimating a global model formulation ( $R = 1$ ) and include dummies for the remaining countries. Thus, estimated country fixed are interpreted as intercept shifts relative to the U.S. When estimating a regional model formulation ( $R = 5$ ), we must account for the fact that we are in a sense estimating five neural networks simultaneously. We include dummies for all but one country within each region separately. We use the U.S. as reference for OECD, Russia as reference for REF, China as reference for Asia, South Africa as reference for MAF, and Mexico as reference for LAM.

Following the standard in the neural networks literature, we minimize the loss function in (2.12) using gradient descent. Since the optimization problem is high-dimensional and nonconvex, the loss function most likely features numerous critical points. As discussed by Goodfellow et al. (2016), second-order methods tend to get stuck in a critical point, even if associated with a high loss value. By contrast, gradient descent is often able to escape critical points associated with a high loss value. However, standard gradient descent performs poorly whenever the Hessian matrix has a poor condition number. It does not know to prefer directions of the parameter space where the slope of the loss function remains negative for longer. Poor conditioning of the Hessian also complicates setting an appropriate step size.<sup>17</sup> One must try to balance the goals of avoiding to overshoot the minimum and being able to make significant progress in directions with little curvature.

We use a popular modification of standard gradient descent, referred to as Adam,<sup>18</sup> which seeks to mitigate the aforementioned issues. It does so through use of momentum and a separate learning rate for each parameter that is automatically adapted at each iteration<sup>19</sup> (Kingma and Ba, 2014). The Adam algorithm is illustrated in Figure B.5 of the appendix using pseudocode. We use suggested defaults for the hyperparameters of the Adam algorithm (Kingma and Ba, 2014). They are found to work well

<sup>17</sup>Often referred to as *learning rate*.

<sup>18</sup>The name Adam is derived from the phrase “ADaptive Moment estimation”.

<sup>19</sup>Often referred to as an *epoch*.

in our experiments. We stop the algorithm once we have seen no significant decrease in the loss function over 100 iterations, using a tolerance level of  $10^{-4}$ . We restore parameter estimates associated with the lowest loss observed across iterations, not necessarily the parameters from the last iteration. For each iteration of the Adam algorithm, we use the entire data set when evaluating the gradient.<sup>20</sup> Note also that cell states are reset between iterations. Since Adam is a local optimization routine, it requires initialization within a well-behaved region of parameter space that is connected to a satisfactory solution by a path the algorithm can follow. We initialize all intercepts, fixed effects, and cell states from zero. This is common practice in the literature [Goodfellow et al., 2016]. We add 1 to the intercept of the forget gate at initialization, as suggested by Jozefowicz, Zaremba, and Sutskever [2015]. Weight matrices applied to the current output from the preceding layer are initialized from a uniform distribution with limits as suggested by Glorot and Bengio [2010], and weight matrices applied to the lagged output from the present layer are initialized using an orthogonal matrix obtained from the QR decomposition of a matrix of random numbers drawn from a normal distribution, as suggested by Saxe, McClelland, and Ganguli [2014]. Slope coefficients in the output layer are initialized from a truncated normal distribution, as suggested by He et al. [2015]. We use multiple initializations to ensure a satisfactory minimum.

### 2.4.3 Model selection

The choice of neural network architecture is a standard model selection task. There is a tradeoff between choosing large values for the depth and width parameters to reduce bias, and choosing small values to ensure smoothness. Within the neural networks literature, hypothesis tests, cross-validation, and information criteria have been proposed for model selection [Anders and Korn, 1999]. Following Bennedsen et al. [2021], we use the Bayesian information criterion (BIC) to decide on optimal network architecture:

$$\text{BIC} = \log J(\hat{\phi}, \hat{\kappa}, \hat{\Gamma}, \hat{\alpha}) + \frac{m \log n}{n}, \quad (2.13)$$

where  $J(\hat{\phi}, \hat{\kappa}, \hat{\Gamma}, \hat{\alpha})$  is the objective function in (2.12) evaluated in parameters estimates,  $n$  is the total number of observations, and  $m$  is the total number of model parameters excluding fixed effects.

We define the set of candidate network architectures following Bennedsen et al. [2021]. The set contains 55 rectangular and pyramid-shaped architectures, as suggested by Masters [1993]. Such architectures are useful for learning gradually more abstract transformations of the inputs and for keeping the number of free parameters at a reasonable level. As discussed by Gu et al. [2020], it is not necessary to include unreasonably many network architectures in the set of candidates. The full

---

<sup>20</sup>Often referred to as *batch learning*.

set of candidate architectures is presented in Table B.4 of the appendix together with the number of free model parameters (excluding fixed effects) associated with each architecture.

## 2.5 Out-of-sample experiment

In this section, we investigate the properties of the path dependent neural network model in a pseudo out-of-sample setting. We find the model is more accurate at predicting global and regional levels of emissions than popular reduced form alternatives from the EKC literature: a quadratic, a cubic, and a spline-based model. We use this as motivation that the path-dependent neural network model is suitable for making scenario-based emissions projections in the next section. We also use this section to motivate the choice of model formulation to be used for making scenario-based emissions projections in the next section (the choice of  $R$ ).

We split the data set described in Section 2.3 into an estimation sample from 1960 to 2013 and a test sample from 2014 to 2018. The estimation sample is used to estimate the models once. Once the models have been estimated, we predict the level of emissions in the test sample conditional on GDP and population growth. When constructing predictions from the path-dependent neural network model, we build memory (seed the cell states of the LSTM layers) for each country separately, using a joint sample where data from the estimation sample is pre-appended to the test data, then keep the predictions for the test sample. For simplicity, we ignore Jensen's inequality, and construct conditional predictions using the relation

$$\hat{Y}_t = \sum_{r=1}^R \sum_{i \in I_r} e^{\hat{y}_{it}(x_{it})} \times \text{POP}_{it}, \quad t = 2014, 2015, \dots, 2018, \quad (2.14)$$

where  $I_r \subseteq \{1, 2, \dots, N\}$  is the set of indices of countries belonging to region  $r$ ,  $R$  is the total number of regions,  $\hat{y}_{it}$  is model output (a prediction of the logarithmic transformation of per capita emissions), and  $\text{POP}_{it}$  is the population size of country  $i$  at time  $t$ .

We consider a global and a regional formulation of the path-dependent neural network model, using the region definitions from Section 2.3 OECD, REF, Asia, MAF, and LAM. We also consider three benchmark models from the EKC literature:

- The quadratic EKC model of Holtz-Eakin and Selden (1995): This model assumes  $y_{it} = \alpha_i + \beta_t + \delta_1 x_{it} + \delta_2 x_{it}^2 + \nu_{i,t}$ , where  $\nu_{i,t}$  is a stochastic error term.
- The cubic EKC model of Grossman and Krueger (1991): This model assumes  $y_{it} = \alpha_i + \beta_t + \delta_1 x_{it} + \delta_2 x_{it}^2 + \delta_3 x_{it}^3 + \nu_{i,t}$ , where  $\nu_{i,t}$  is a stochastic error term.
- The spline-based EKC model of Schmalensee et al. (1998): This model assumes  $y_{it} = \alpha_i + \beta_t + f^{\text{splines}}(x_{it}) + \nu_{i,t}$ , where  $f^{\text{splines}}(\cdot)$  denotes 10-segment piecewise linear splines, and  $\nu_{i,t}$  is a stochastic error term.

**Table 2.3:** Optimal network architectures for the path-dependent neural network model

	Global	Regional
Neural network architecture	(4, 4, 4)	(2, 2, 2)
# parameters (excl. fixed effects)	388	122

Note: "( $a,b,c$ )" indicates a neural network architecture with three hidden layers containing  $a$  units in the first layer,  $b$  in the second, and  $c$  in the third.

We estimate a global formulation of the benchmark models, where it is assumed that the shape of the income-emissions relationship is common to all countries, and we also estimate the benchmark models region-wise to parallel the regional formulation of the path-dependent neural network model. However, note that region-wise estimation of the benchmark models ignores cross-region dependencies. By contrast, the regional formulation of the path-dependent neural network model uses information from across regions to learn the parameters of the hidden layers of the model, but uses within-region information to learn the parameters of the output layer. Recall from the discussion surrounding Table 2.1 that it is likely important to be able to use information from across regions when making scenario-based emissions projections based on the SSPs, as values of per capita GDP from the SSPs are often out of sample range for some regions, but within sample range for other regions. Such issues are less severe when looking only five years into the future. The fraction of observations on per capita GDP in the test sample within range of the estimation sample is 1.00, 0.92, 1.00, 0.98, and 0.92 for OECD, REF, Asia, MAF, and LAM, respectively. This out-of-sample experiment is therefore likely overly optimistic about the usefulness of the regional formulation of the benchmark models for making scenario-based emissions projections.

The benchmark models do not have readily available time fixed effects for the test sample. As mentioned in the previous section, Schmalensee et al. (1998) suggest to project time fixed effects into the future by fitting a simple parametric function to the time fixed effects for the estimation sample, then use that function to extrapolate out-of-sample. Figure B.3 of the appendix shows the estimated time fixed effects from the quadratic EKC model. It is not obvious to us what would be appropriate functions to fit to these series. Estimated time fixed effects for the remaining benchmark models look similar to those in Figure B.3. To avoid having out-of-sample performance driven by an arbitrary choice of function, we follow the approach suggested by Holtz-Eakin and Selden (1995), and use the estimated time fixed effects for the last period of the estimation sample when constructing predictions from the benchmark models. Both the approaches of Holtz-Eakin and Selden (1995) and Schmalensee et al. (1998) have obvious limitations that seem to increase the further we look ahead in time. As this out-of-sample experiment looks only five years ahead, it is likely overly optimistic

**Table 2.4:** Out-of-sample relative predictive accuracy

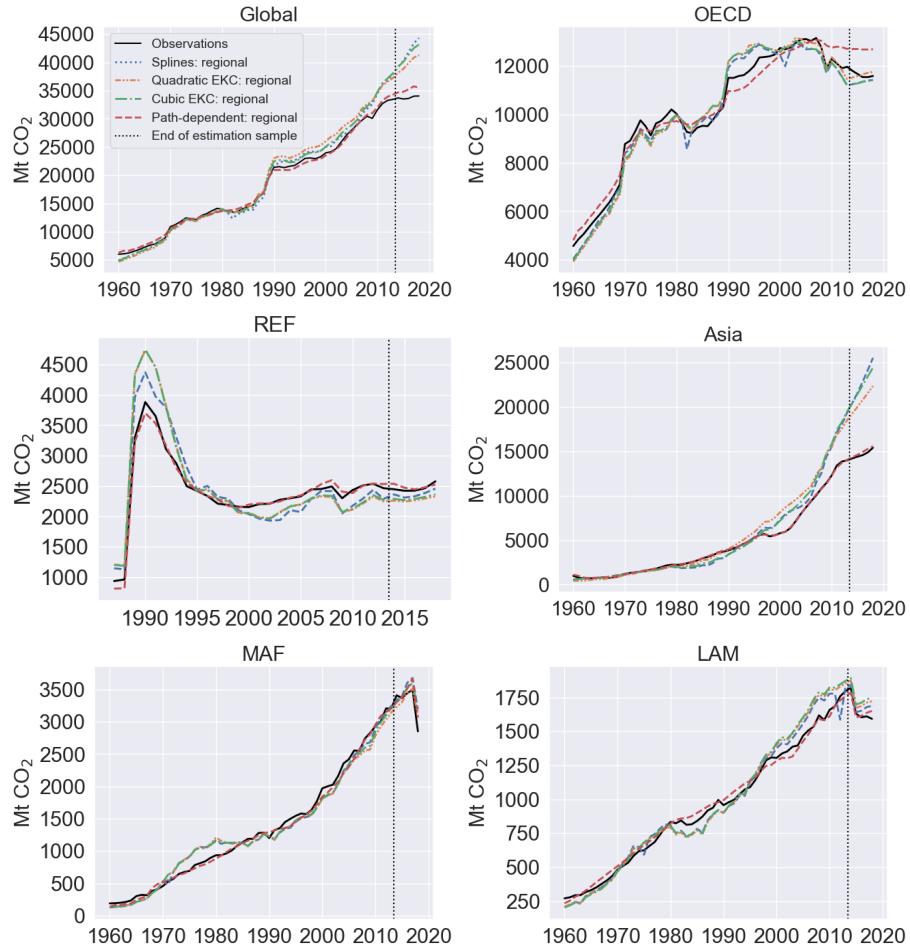
	Global	OECD	REF	Asia	MAF	LAM
<b>Quadratic EKC model</b>						
Global	1.000	1.000	1.000	1.000	1.000	1.000
Regional	0.136	<b>0.001</b>	0.067	0.307	0.106	2.767
<b>Cubic EKC model</b>						
Global	0.895	0.572	1.317	1.011	0.497	0.610
Regional	0.210	0.004	0.049	0.497	0.153	4.436
<b>Spline-based EKC model</b>						
Global	0.955	0.316	1.647	1.264	0.578	6.810
Regional	0.245	0.004	0.018	0.565	0.216	1.199
<b>Path-dependent model</b>						
Global	0.067	0.019	0.196	0.092	0.150	<b>0.101</b>
Regional	<b>0.008</b>	0.043	<b>0.005</b>	<b>0.001</b>	<b>0.105</b>	0.445

Note: Numbers reflect relative mean squared errors over the test sample, using the global formulation of the quadratic EKC model as benchmark. Boldface numbers indicate models with the lowest MSE for each region. The regional formulation of the path-dependent neural network model is highlighted as the preferred mode overall.

about the usefulness of the benchmark models (global and regional formulation) for making long-run scenario-based emissions projections.

We report optimal network architectures for the path-dependent neural network model in Table 2.3 obtained by minimizing the BIC in 2.13. For both the global and the regional model formulation, the bias-variance tradeoff favors a deep, but narrow and rectangular architecture, containing three hidden layers with 4 and 2 units in each, respectively. Although the two architectures are the 10th and 4th simplest in the candidate set in terms of number of parameters, they contain 388 and 122 parameters excluding fixed effects. One might think that such large numbers of free parameters would lead to overfitting. As we shall see below, this is not necessarily the case. The models require a lot of parameters, because the parameters are used for distinct purposes in the different components of the LSTM memory cells; recall the discussion surrounding equations 2.7 - 2.11 in Section 2.4.

Table 2.4 reports relative predictive accuracy for the models considered over the test. Numbers in the table reflect relative mean squared errors (MSEs), using the global formulation of the quadratic EKC model as benchmark. The first column of numbers shows relative MSEs for the global level of emissions and is our main focus. To better understand global accuracy, remaining columns show relative MSEs for each region separately. Conditional predictions for each region are obtained from

**Figure 2.5:** Out-of-sample predictions: Regional model formulations

the second sum in 2.14. The path-dependent neural network model produces the most accurate predictions for four of the five regions considered, leading to the most accurate predictions for global emissions. The quadratic EKC model produces the most accurate predictions for OECD, and the second-most accurate predictions for global emissions. The cubic EKC model produces the third-most accurate predictions for global emissions, and the spline-based EKC model produces less accurate predictions for global emissions than the other models. Note also that the differences in accuracy for global emissions between the benchmark models are smaller than the differences in accuracy between the benchmark models and the path-dependent neural network model.

Inspecting Table 2.4, we note the bias-variance tradeoff favors the regional model

formulation to the global model formulation. For each model, the regional formulation produces the most accurate predictions for four of the five regions considered, leading to the most accurate predictions for global emissions. This is interesting since the literature has focused almost exclusively on global model formulations [Holtz-Eakin and Selden, 1995] [Schmalensee et al., 1998]. The regional formulation of the path-dependent neural network model is the preferred model overall. Of all models considered, it produces the most accurate predictions for three of the five regions, leading to the most accurate predictions for global emissions. We, therefore, use the regional formulation of the path-dependent neural network model to construct scenario-based emissions projections in the next section.

Figure 2.5 plots the predictions from the regional formulation of each model to get a better understanding of how and where errors occur. We note the largest errors occur for Asia. All benchmark models severely overpredict the level of emissions for Asia, starting already within the estimation sample. Of the benchmark models, the quadratic EKC model overpredicts the level of emissions for Asia slightly less so than the cubic and the spline-based EKC model. However, only the path-dependent neural network model is able to capture the trend breaks observed for Asia toward the end of the 1990s and the beginning of the 2010s. The large differences in predictive accuracy between the benchmark models and the path-dependent neural network model observed for Asia seem to be the main driver of the differences observed for global emissions. For OECD, the benchmark models are slightly more accurate than the path-dependent neural network model, which is overpredicting the level of emissions for OECD. For the remaining regions, the path dependent neural network model is outperforming the benchmark models by a small margin. Figure B.2 of the appendix plots the predictions from the global formulation of each model.

## 2.6 SSP projections

In this section, we construct CO<sub>2</sub> emissions projections through 2100, using the regional formulation of the path-dependent neural network model and national-level scenarios for per capita GDP from the SSP framework discussed in Section 2.3.1. We compare our reduced-form model projections to baseline projections (SSP baseline scenarios) and mitigation projections (SSP mitigation scenarios) from the structural IAMs IMAGE, MESSAGE-GLOBIOM, AIM/CGE, GCAM4, REMIND-MAGPIE, and WITCH-GLOBIOM, available through the SSP database<sup>21</sup> to assess whether the IAM projections appear consistent with historical experience.

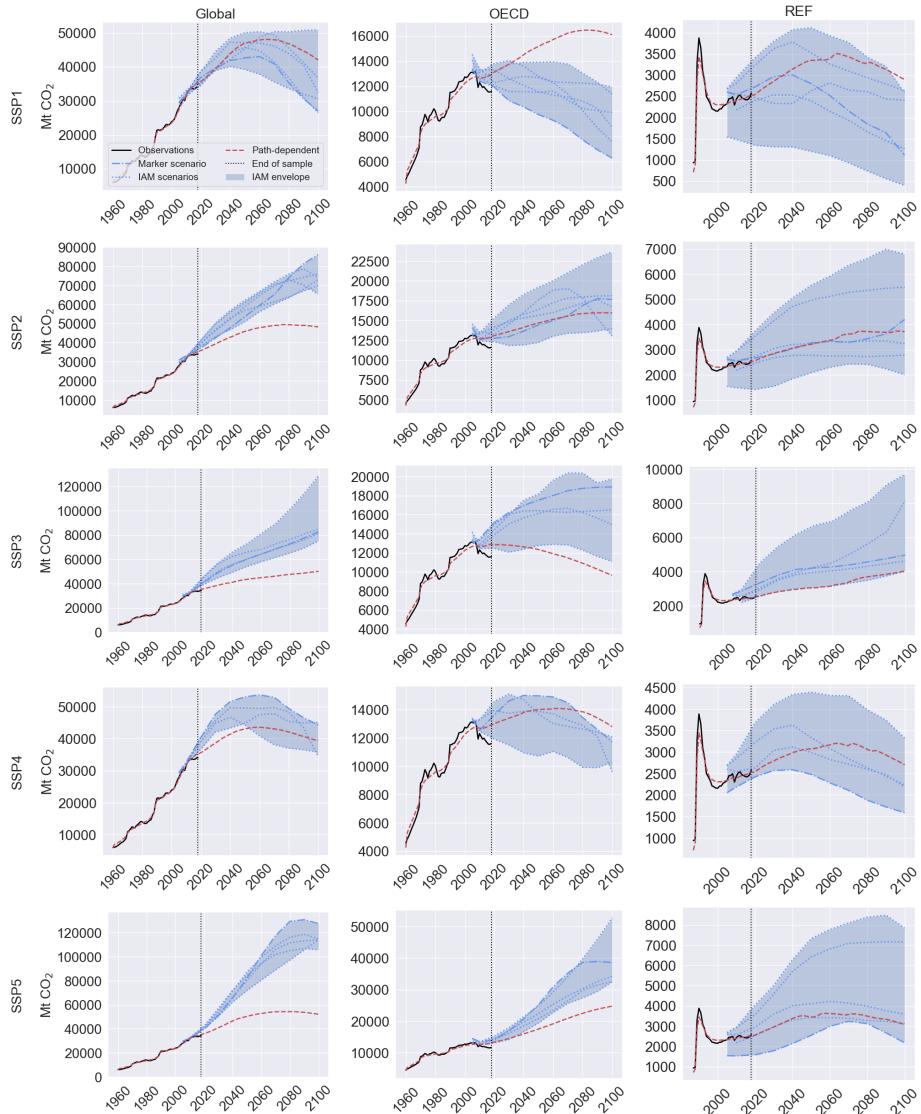
We estimate the regional formulation of the path-dependent neural network model, using the region definitions and the data set described in Section 2.3. We construct future CO<sub>2</sub> emissions projections through 2100, using the national-level

<sup>21</sup> Accessible at [tntcat.iiasa.ac.at/SspDB](http://tntcat.iiasa.ac.at/SspDB). IAM baseline scenarios for CO<sub>2</sub> emissions from fossil fuels and industry (unharmonized) was downloaded on July 9, 2021. IAM mitigation scenarios for CO<sub>2</sub> emissions from fossil fuels and industry (unharmonized) was downloaded on August 10, 2021.

scenarios for per capita GDP from SSP1-SSP5 (SSP main scenarios), described in Section 2.3.1. There are subtle differences between the sample of countries in the data set used for estimation, and the sample of countries in the SSPs. We indicate such differences in Table B.1 of the appendix. The estimation sample contains 186 countries (names in black and names in green in Table B.1). The SSPs contain 184 countries (names in black and names in red in Table B.1). The intersection of the two sets of countries contains 174 countries (names in black in Table B.1), which we use to construct emissions projections. The optimal network architecture for the regional formulation of the path-dependent neural network model contains three hidden layers with four units in each layer. This is a slightly more complex architecture than in the out-of-sample experiment of the previous section, where the estimation sample was cut off by 2013. To ensure robustness, we estimate the model ten times using different initializations, then average the projections. For each SSP, values of per capita GDP are available at a five-year frequency. Since the path-dependent neural network model is estimated using yearly data, memory cells of the LSTM layers expect data to be yearly. Therefore, we linearly interpolate values between the five-year data. When constructing emissions projections, we build memory in the model (seed the cell states of the LSTM layers) for each country separately using a joint sample where the estimation data is pre-appended to the scenario-data, then keep the scenario projections.

In Figures 2.6 and 2.7 we plot projections from our path-dependent neural network model together with baseline IAM projections. As we discuss in Section 2.3.1 baseline IAM projections are meant to reflect future developments in absence of new climate policies. The IAM projections date back to 2005. Although the IAM projections are constructed after 2005, they differ somewhat for the initial time periods. This is because the IAMs rely on different inventories for base year calibration (van Vuuren et al., 2011). It is not clear at precisely what point in time projections from the different IAMs are constructed. Initially, we may note from Figures 2.6 and 2.7 that our emissions projections seem to reflect socioeconomic challenges for mitigation, as described by Figure 2.1; we project the highest level of global emissions by 2100 for the scenarios characterized by the highest level of socioeconomic challenges for mitigation SSP3 and SSP5; we project the lowest level of global emissions by 2100 for the scenarios characterized by the lowest level of socioeconomic challenges for mitigation SSP1 and SSP4; and, we project an intermediate level of global emissions by 2100 for the middle of the road scenario SSP2. Our reduced-form emission projections reflect historical records and provide a benchmark against which the structural IAM projections can be compared. For each country, our model projections are dependent on the path for per capita GDP that country is on and factors implicitly associated with that path are assumed to continue in the future. By contrast, IAM projections include structural shifts and vary more across SSPs than our emissions projections, as they include more factors explicitly.

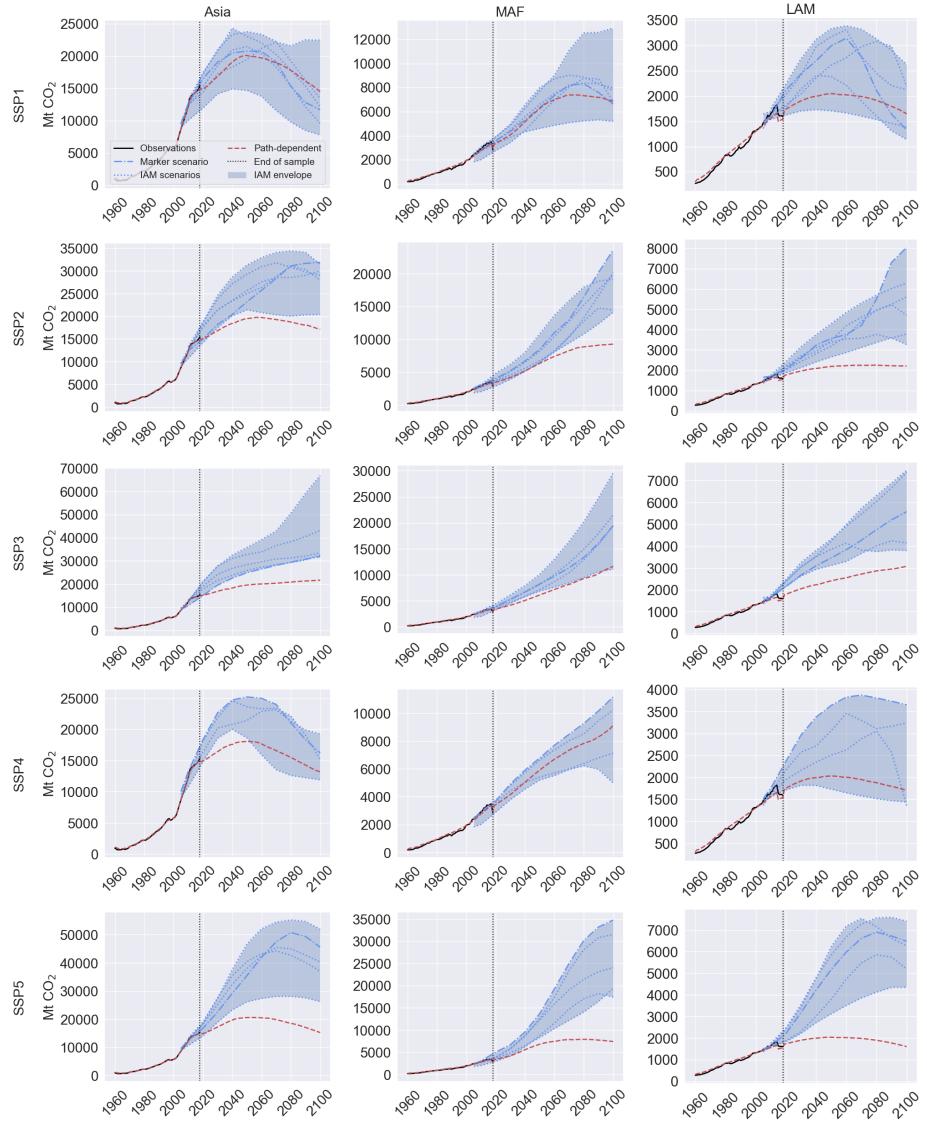
**Figure 2.6:** Path-dependent neural network model projections versus baseline IAM projections: global, OECD, and REF



Note: Baseline IAM projections are from IMAGE, MESSAGE-GLOBIOM, AIM/CGE, GCAM4, REMIND-MAGPIE, and WITCH-GLOBIOM. Projections from all IAMs are not available for every SSP. Projections from the path-dependent neural network model are based on a regional model formulation.

For the scenarios with lowest socioeconomic challenges for mitigation SSP1 and SSP4, our global emissions projections are within the envelope of baseline projections

**Figure 2.7:** Path-dependent neural network model projections versus baseline IAM projections: Asia, MAF and LAM



Note: Baseline IAM projections are from IMAGE, MESSAGE-GLOBIOM, AIM/CGE, GCAM4, REMIND-MAGPIE, and WITCH-GLOBIOM. Projections from all IAMs are not available for every SSP. Projections from the path-dependent neural network model are based on a regional model formulation.

from the IAMs for most years. The same is true for most regions separately. We interpret this as indication the baseline IAM projections for SSP1 and SSP4 seem

consistent with historical experience. The only exception is for OECD and SSP1. Here, we project systematically higher emissions than the IAMs, and a much later turning point around 2080. Similar to the IAMs, we predict global emissions to reach a turning point slightly after 2060 for SSP1, and little prior to 2060 for SSP4, and we predict emissions to reach a turning point within most regions separately.

Recall from the discussion in Section 2.3 that SSP2 is meant to reflect a world in which social, economic, and technological trends do not shift markedly from historical trends. Yet, we predict global emissions to follow a trajectory below the baseline IAMs projections. This could indicate that our reduced-form model is more optimistic about effects of current climate policies than the baseline IAM projections. In our model, effects of climate policies are captured indirectly through their effect on per capita GDP. For SSP2, we predict a turning point in global emissions around 2080 that seems to be mainly driven by a turning point for Asia. The difference between our global emissions projections and those of the IAMs appears to stem mostly from the regions Asia, MAF, and LAM.

For the scenarios characterized by the highest socioeconomic challenges for mitigation SSP3 and SSP5, our global emissions projections are the farthest below the baseline IAM projections. SSP3, characterized by high socioeconomic challenges for both mitigation and adaptation, is the only scenario for which we predict no turning point in global emissions and neither do the IAMs. Across regions, our emissions projections for SSP3 are toward the bottom or slightly below the envelope of baseline IAM projections. Similar to the IAMs, we predict emissions reach to a turning point prior 2100 only within OECD. Among many other factors, SSP3 characterizes a world described by increasingly low international priority for addressing environmental concerns (Riahi et al. 2017). This is expected to lead to strong environmental degradation in some regions, and may help to explain why our reduced-form model projects lower levels of emissions than the baseline IAM projections. For SSP5, we also predict global emissions to follow a trajectory below the baseline IAM projections. By 2100, we predict that emissions have reached only half (or less) the level of the IAM projections. Only for the region REF our emissions projections are within the envelope of baseline IAM projections. Among many other factors, SSP5 characterizes a world described by adoption of resource and energy intensive lifestyles and exploitation of abundant fossil fuel resources (Riahi et al. 2017). This may help to explain why our reduced-form model projects lower levels of emissions than the baseline IAM projections.

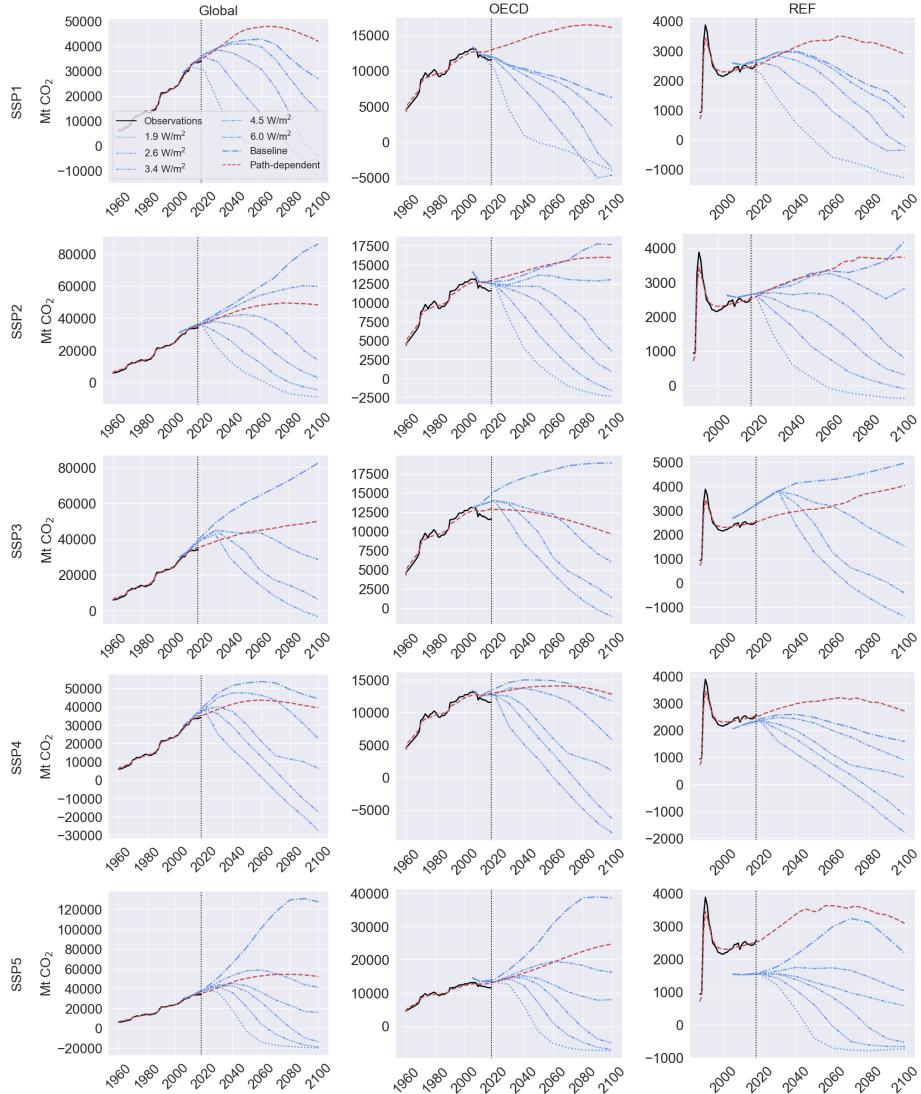
By the end of the sample period (2018), Asia is emitting the highest level of CO<sub>2</sub> followed by OECD. For most scenarios, we predict these regions will remain the largest CO<sub>2</sub> emitters through 2100. However, for these regions, we predict early turning points in emissions in most scenarios. SSP3 is the only scenario for which we do not predict a turning point for Asia, SSP5 and SSP2 are the only scenarios for which we do not predict a clear turning point in emissions for OECD. By contrast, we

predict strictly increasing emissions for MAF in most scenarios. Similar to the IAMs, we therefore predict MAF will eventually overtake OECD in terms of emissions in SSP3. Even for some other scenarios, such as SSP4, it seems MAF would become a (or the) top emitter soon after 2100 if trends observed in Figures 2.6 and 2.7 continue.

In Figures 2.8 and 2.9 we plot our emissions projections from the path-dependent neural network model together with IAM mitigation projections from the marker models discussed in Section 2.3.1 which consider implications for emissions of climate change mitigation policies. For each SSP, our model projections are plotted together with IAM mitigation projections with different target forcing levels for 2100. Especially for SSP3 and SSP5 and the region REF, the IAM mitigation projections appear misaligned with historical observations. Recall that the IAMs are calibrated on different inventories, and they are not calibrated specifically on historical levels of CO<sub>2</sub> emissions for the five regions defined for the SSPs (Riahi et al. 2017). Therefore, the marker models for SSP3 and SSP5, AIM-CGE and REMIND-MAGPIE, respectively, are not guaranteed to provide an accurate fit for CO<sub>2</sub> emissions for a small region such as REF even though they provide an accurate global fit. For the scenarios with lowest socioeconomic challenges for mitigation SSP1 and SSP4, we discussed above that our global emissions projections seemed consistent with the baseline IAM projections. This is also reflected by Figures 2.8 and 2.9. For the scenarios with medium socioeconomic challenges for mitigation (SSP2) and high socioeconomic challenges for mitigation (SSP3 and SSP5), we discussed above that our global emissions projections follow a trajectory below the baseline IAM projections. We note from Figures 2.8 and 2.9 that our global emissions projections for these scenarios follow a trajectory closer to the trajectory from IAM mitigation projections with a target forcing of 6.0 W/m<sup>2</sup> than remaining IAM baseline and mitigation projections. We interpret this as indication that the IAM projections with a target forcing of 6.0 W/m<sup>2</sup> appear more consistent with historical experience than remaining baseline and mitigation projections for SSP2, SSP3, and SSP5. This could indicate that climate policies associated with IAM projections that target a forcing of 6.0 W/m<sup>2</sup> are more consistent with historical experience than those associated with remaining baseline and mitigation projections for these scenarios.

For SSP2, our global emissions projections follow a trajectory slightly below the trajectory from IAM projections with a target forcing of 6.0 W/m<sup>2</sup>, but above the trajectory from IAM projections with a target forcing of 4.5 W/m<sup>2</sup>. Our projections seem pulled below the 6.0 W/m<sup>2</sup> trajectory by Asia, MAF, and LAM. The projections for OECD and REF seem close to the baseline IAM projections (as in Figure 2.6). For SSP3, our global emissions projections start slightly below the trajectory from IAM projections with a target forcing of 6.0 W/m<sup>2</sup>, then intersect the 6.0 W/m<sup>2</sup> trajectory around 2050, and remain above it through 2100. The same is true for most regions separately, albeit with slight varying years of intersection. For SSP5, our global emissions projections also start slightly below the trajectory from IAM projections

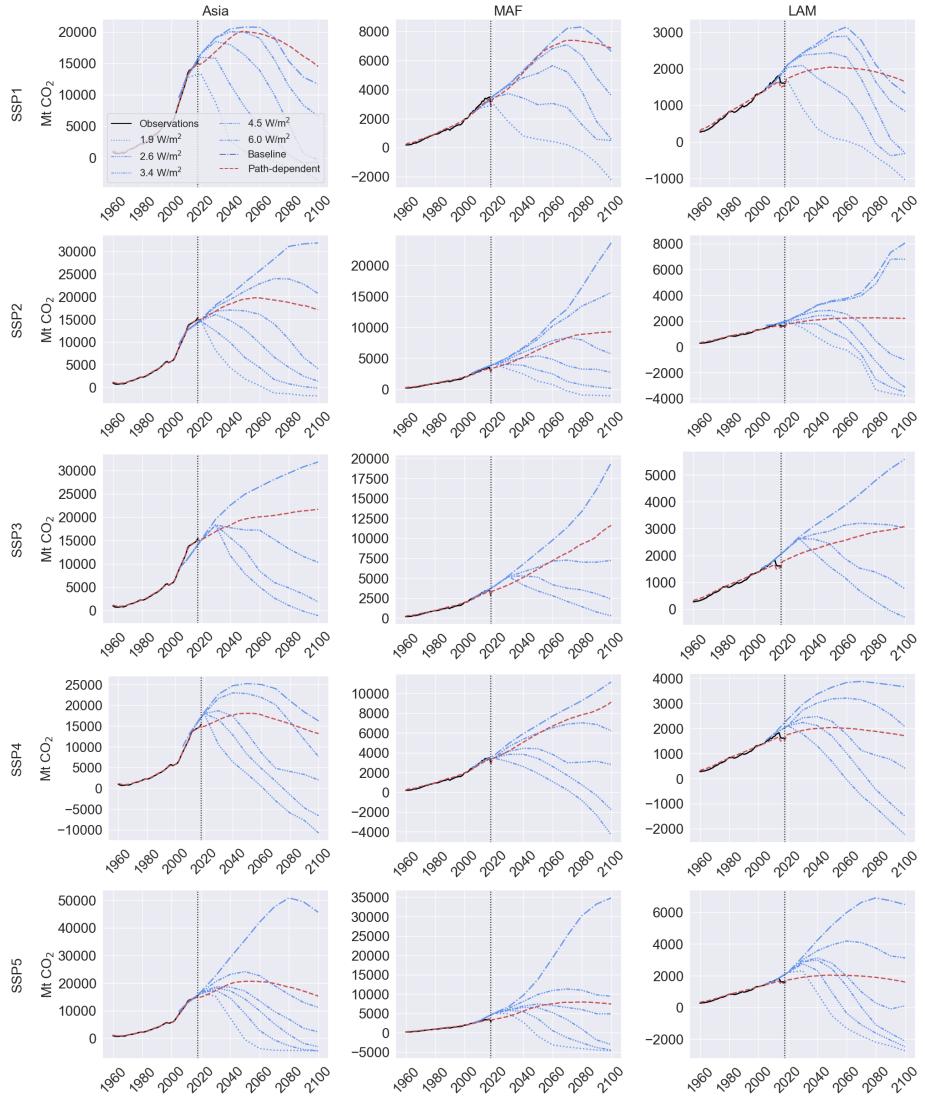
**Figure 2.8:** Path-dependent neural network model projections versus IAM mitigation projections: global, OECD, and REF



Note: For SSP1-SSP5, mitigation projections are from the markers IMAGE, MESSAGE-GLOBIOM, AIM/CGE, GCAM4, and REMIND-MAGPIE, respectively. Not all forcing targets are reachable for every SSP. Net negative levels of CO<sub>2</sub> emissions occur when anthropogenic removals exceed anthropogenic emissions. Projections from the path-dependent neural network model are based on a regional model formulation.

with a target forcing of  $6.0 \text{ W/m}^2$ , then intersect the  $6.0 \text{ W/m}^2$  trajectory around

**Figure 2.9:** Path-dependent neural network model projections versus IAM mitigation projections: Asia, MAF, and LAM



Note: For SSP1-SSP5, mitigation projections are from the markers IMAGE, MESSAGE-GLOBIOM, AIM/CGE, GCAM4, and REMIND-MAGPIE, respectively. Not all forcing targets are reachable for every SSP. Net negative levels of CO<sub>2</sub> emissions occur when anthropogenic removals exceed anthropogenic emissions. Projections from the path-dependent neural network model are based on a regional model formulation.

2070, and remain above it through 2100. For SSP5, our global emissions projections

seem pulled above the  $6.0 \text{ W/m}^2$  trajectory after 2070 especially by OECD and Asia.

As our emissions projections for all regions and for all scenarios are above the trajectory from IAM projections with a target forcing of  $1.9 \text{ W/m}^2$  by 2100, we predict large-scale structural changes are necessary to reach the goals of the Paris Agreement, which are broadly in line with trajectories targeting a forcing level of  $1.9 \text{ W/m}^2$  by 2100.

## 2.7 Conclusion

We extend the semiparametric panel data methodology developed in [Bennedsen et al. 2021], and propose a novel neural network-based panel data model for constructing scenario-based carbon dioxide ( $\text{CO}_2$ ) emissions projections that can be readily used out-of-sample by avoiding explicit dependence on time. The model consists of additively separable country fixed effects and a long short-term memory (LSTM) recurrent neural network regression component. The LSTM network regression component takes into account time implicitly by building memory and letting model predictions depend on the income path a country is on. In a pseudo out-of-sample experiment, we demonstrate the model more accurately predicts global and regional levels of emissions than popular alternatives from the environmental Kuznets curve literature [Grossman and Krueger 1991; Holtz-Eakin and Selden 1995; Schmalensee et al. 1998].

We project  $\text{CO}_2$  emissions through 2100, using national-level scenarios for per capita gross domestic product from the Shared Socioeconomic Pathways (SSPs; Riahi et al. 2017). By 2100, we project the lowest level of global emissions for scenarios with the lowest socioeconomic challenges for mitigation SSP1 and SSP4. For these scenarios, our projections appear consistent with baseline projections from structural integrated assessment models (IAMs) that are meant to characterize future developments in absence of new climate policies beyond those already in place today. We project the highest level of global  $\text{CO}_2$  emissions by 2100 for scenarios with the highest socioeconomic challenges for mitigation SSP3 and SSP5, and we project an intermediate level of emissions for the “middle of the road” scenario SSP2. For these scenarios, our projections appear the most consistent with mitigation projections from IAMs that target a forcing level of  $6.0 \text{ W/m}^2$  by 2100 through climate change mitigation policies. For all scenarios and for all regions, our emissions projections are above mitigation projections from IAMs that target a forcing level of  $1.9 \text{ W/m}^2$  by 2100. We therefore predict large-scale structural changes are required to reach the goals of the Paris Agreement, which is broadly in line with mitigation scenarios that target a forcing level of  $1.9 \text{ W/m}^2$  by 2100 (Riahi et al. 2017).

There are many directions in which our work could be extended in the future, and we believe our work could be useful in a range of different settings, such as for prediction of panels of economic data. The model proposed in this paper can be

readily used with multiple input variables, and since neural network models often perform well even in high-dimensional settings with a low signal-to-noise ratio [Gu et al. 2020; Christensen, Siggaard, and Veliyev 2021; Jensen 2021], our work may also be useful for modeling high-dimensional and high-frequency financial data.

### Acknowledgements

For useful comments and suggestions, we thank participants at EGU2020: Sharing Geoscience online (2020) and the Joint Econometrics-Finance seminar (2020) at the Center for Research in Econometric Analysis of Time Series (CREATES) at Aarhus University. The authors also acknowledge financial support from the DFF grant Econometric Modeling of Climate Change.

## 2.8 References

- Alcamo et al., 1995. An Evaluation of the IPCC IS92 Emission Scenarios. Iiasa research report (reprint), IIASA, Laxenburg, Austria.
- URL <http://pure.iiasa.ac.at/id/eprint/4465/>
- Anders, U., Korn, O., 1999. Model selection in neural networks. *Neural Networks* 12 (2), 309–323.
- URL <http://www.sciencedirect.com/science/article/pii/S0893608098001178>
- Auffhammer, M., Steinhauser, R., 2012. Forecasting The Path of U.S. CO<sub>2</sub> Emissions Using State-Level Information. *The Review of Economics and Statistics* 94 (1), 172–185.
- URL [https://doi.org/10.1162/REST\\_a\\_00152](https://doi.org/10.1162/REST_a_00152)
- Ausubel, J. H., Nordhaus, W. D., 1983. A review of estimates of future carbon dioxide emissions. *Changing Climate*, 153–185.
- Bengio, Y., Frasconi, P., Simard, P., 1993. The problem of learning long-term dependencies in recurrent networks. In: IEEE International Conference on Neural Networks. 1183–1188 vol.3.
- Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* 5 (2), 157–166.
- Bennedsen, M., Hillebrand, E., Jensen, S., 2021. A Neural Network Approach to the Environmental Kuznets Curve, Chapter 1 of this dissertation.
- Bouwman, A., Kram, T., Klein Goldewijk, K., 2006. Integrated modelling of global environmental change: an overview of IMAGE 2.4. Netherlands Environmental Assessment Agency (MNP), Bilthoven, The Netherlands.
- Calvin, K., 2011. The global Change Assement Model. <http://wiki.umd.edu/gcam/>.
- Calvin et al., 2017. The SSP4: A world of deepening inequality. *Global Environmental Change* 42, 284–296.
- URL <http://www.sciencedirect.com/science/article/pii/S095937801630084X>
- Chang, Y., Choi, Y., Kim, C. S., Miller, J. I., Park, J. Y., 2016. Disentangling temporal patterns in elasticities: A functional coefficient panel analysis of electricity demand. *Energy Economics* 60, 232–243.
- URL <https://www.sciencedirect.com/science/article/pii/S0140988316302717>
- Chang, Y., Choi, Y., Kim, C. S., Miller, J. I., Park, J. Y., 2021. Forecasting regional long-run energy demand: A functional coefficient panel approach. *Energy Economics* 96, 105117.
- URL <https://www.sciencedirect.com/science/article/pii/S0140988321000220>

- Christensen, K., Siggaard, M., Veliyev, B., Jan. 2021. A machine learning approach to volatility forecasting. WorkingPaper 2021-03, Institut for Økonomi, Aarhus Universitet.
- Cybenko, G., Dec 1989. Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals and Systems 2 (4), 303–314.  
 URL <https://doi.org/10.1007/BF02551274>
- de Bruyn, S., van den Bergh, J., Opschoor, J., 1998. Economic growth and emissions: reconsidering the empirical basis of environmental Kuznets curves. Ecological Economics 25 (2), 161 – 175.  
 URL <http://www.sciencedirect.com/science/article/pii/S092180099700178X>
- Dellink, R., Chateau, J., Lanzi, E., Magné, B., 2017. Long-term economic growth projections in the Shared Socioeconomic Pathways. Global Environmental Change 42, 200–214.  
 URL <https://www.sciencedirect.com/science/article/pii/S0959378015000837>
- Dijkgraaf, E., Vollebergh, H. R. J., Oct. 2005. A Test for Parameter Homogeneity in CO2 Panel EKC Estimations. Environmental and Resource Economics 32 (2), 229—239.  
 URL <https://doi.org/10.1007/s10640-005-2776-0>
- Emmerling et al., 2016. The WITCH 2016 Model - Documentation and Implementation of the Shared Socioeconomic Pathways. Tech. rep., Fondazione Eni Enrico Mattei (FEEM).  
 URL <http://www.jstor.org/stable/resrep15162.4>
- Fricko et al., 2017. The marker quantification of the Shared Socioeconomic Pathway 2: A middle-of-the-road scenario for the 21st century. Global Environmental Change 42, 251 – 267.  
 URL <http://www.sciencedirect.com/science/article/pii/S0959378016300784>
- Fujimori, S., Masui, T., Yuzuru, M., 2012. AIM/CGE [basic] manual. Discussion paper series. Tsukuba, Japan: Center for Social and Environmental Systems Research, NIES.
- Fujimori et al., 2017. SSP3: AIM implementation of Shared Socioeconomic Pathways. Global Environmental Change 42, 268 – 283.  
 URL <http://www.sciencedirect.com/science/article/pii/S0959378016300838>
- Galeotti, M., Lanza, A., Pauli, F., 2006. Reassessing the environmental Kuznets curve for CO<sub>2</sub> emissions: A robustness exercise. Ecological Economics 57 (1), 152 – 163.  
 URL <http://www.sciencedirect.com/science/article/pii/S0921800905001680>
- Gers, F., Schmidhuber, J., Cummins, F., 1999. Learning to forget: continual prediction with LSTM. In: 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470). Vol. 2. 850–855 vol.2.

- Global Carbon Project, 2019. Supplemental data of Global Carbon Budget 2019 (Version 1.0) [Data set]. Global Carbon Project.
- Glorot, X., Bengio, Y., 13–15 May 2010. Understanding the difficulty of training deep feedforward neural networks. In: Teh, Y. W., Titterington, M. (Eds.), Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Vol. 9 of Proceedings of Machine Learning Research. PMLR, Chia Laguna Resort, Sardinia, Italy, 249–256.  
URL <http://proceedings.mlr.press/v9/glorot10a.html>
- Glorot, X., Bordes, A., Bengio, Y., 11–13 Apr 2011. Deep Sparse Rectifier Neural Networks. In: Gordon, G., Dunson, D., Dudík, M. (Eds.), Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Vol. 15 of Proceedings of Machine Learning Research. PMLR, Fort Lauderdale, FL, USA, 315–323.  
URL <http://proceedings.mlr.press/v15/glorot11a.html>
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. <http://www.deeplearningbook.org>.
- Graves, A., rahman Mohamed, A., Hinton, G., 2013. Speech Recognition with Deep Recurrent Neural Networks.
- Grossman, G. M., Krueger, A. B., November 1991. Environmental Impacts of a North American Free Trade Agreement. Working Paper 3914, National Bureau of Economic Research.  
URL <http://www.nber.org/papers/w3914>
- Gu, S., Kelly, B., Xiu, D., 02 2020. Empirical Asset Pricing via Machine Learning. The Review of Financial Studies 33 (5), 2223–2273.  
URL <https://doi.org/10.1093/rfs/hhaa009>
- Halkos, G. E., Tsionas, E. G., 2001. Environmental Kuznets curves: Bayesian evidence from switching regime models. Energy Economics 23 (2), 191 – 210.  
URL <http://www.sciencedirect.com/science/article/pii/S0140988300000633>
- Harbaugh, W. T., Levinson, A., Wilson, D. M., 2002. Reexamining the Empirical Evidence for an Environmental Kuznets Curve. The Review of Economics and Statistics 84 (3), 541–551.  
URL <http://www.jstor.org/stable/3211570>
- He, K., Zhang, X., Ren, S., Sun, J., December 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: The IEEE International Conference on Computer Vision (ICCV).
- Hendry, D., 2018. First-in, first-out: Driving the UK's per capita carbon dioxide emissions below 1860 levels. VoxEU.

- Hochreiter, S., 1991. Untersuchungen zu dynamischen neuronalen Netzen. Diploma, Technische Universität München 91 (1).
- Holtz-Eakin, D., Selden, T. M., 1995. Stoking the fires? CO<sub>2</sub> emissions and economic growth. *Journal of Public Economics* 57 (1), 85 – 101.  
 URL <http://www.sciencedirect.com/science/article/pii/004727279401449X>
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2 (5), 359–366.  
 URL <https://www.sciencedirect.com/science/article/pii/0893608089900208>
- Hyötyniemi, H., 1996. Turing machines are recurrent neural networks. *Proceedings of step 96*.
- IPCC, 1995. SAR Climate Change 1995: Synthesis Report.
- IPCC, 2001. TAR Climate Change 2001: Synthesis Report.
- IPCC, 2014. AR5 Synthesis Report: Climate Change 2014.
- IPCC, 2021. AR6 Climate Change 2021: The Physical Science Basis.
- Jensen, S., 2021. Nowcasting U.S. CO<sub>2</sub> Emissions using Machine Learning, working paper.
- Jozefowicz, R., Zaremba, W., Sutskever, I., 2015. An Empirical Exploration of Recurrent Network Architectures. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37. ICML'15. JMLR.org, p. 2342–2350.
- Kingma, D. P., Ba, J., 2014. Adam: A Method for Stochastic Optimization.
- Kriegler et al., 2017. Fossil-fueled development (SSP5): An energy and resource intensive scenario for the 21st century. *Global Environmental Change* 42, 297 – 315.  
 URL <http://www.sciencedirect.com/science/article/pii/S0959378016300711>
- Leshno, M., Lin, V. Y., Pinkus, A., Schocken, S., 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks* 6 (6), 861–867.  
 URL <https://www.sciencedirect.com/science/article/pii/S0893608005801315>
- List, J. A., Gallet, C. A., 1999. The environmental Kuznets curve: does one size fit all? *Ecological Economics* 31 (3), 409 – 423.  
 URL <http://www.sciencedirect.com/science/article/pii/S0921800999000646>
- Luderer et al., 2013. Description of the REMIND Model (Version 1.5). Available at SSRN: <https://ssrn.com/abstract=2312844> or <http://dx.doi.org/10.2139/ssrn.2312844>.

- Masters, T., 1993. Practical neural network recipes in C++. (Academic Press).
- Messner, S., Strubegger, M., July 1995. User's Guide for MESSAGE III. IIASA Working Paper, IIASA, Laxenburg, Austria.  
URL <http://pure.iiasa.ac.at/id/eprint/4527/>
- Myhre et al., 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Ch. 8. Anthropogenic and Natural Radiative Forcing, p. 659–740.
- Müller-Fürstenberger, G., Wagner, M., 2007. Exploring the environmental Kuznets hypothesis: Theoretical and econometric problems. Ecological Economics 62 (3), 648 – 660.  
URL <http://www.sciencedirect.com/science/article/pii/S0921800906004058>
- OECD, 2018. Income inequality.  
URL <https://www.oecd-ilibrary.org/content/data/459aa7f1-en>
- O'Neill et al., 2014. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. Climatic Change 122 (3), 387–400.  
URL <https://doi.org/10.1007/s10584-013-0905-2>
- Preeti, Bala, R., Singh, R. P., 2019. Financial and Non-Stationary Time Series Forecasting using LSTM Recurrent Neural Network for Short and Long Horizon. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). 1–7.
- Ramachandran, P., Zoph, B., Le, Q. V., 2017. Searching for Activation Functions.
- Riahi et al., 2017. Supplementray Material: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change 42, 153 – 168.  
URL <http://www.sciencedirect.com/science/article/pii/S0959378016300681>
- Romero-Ávila, D., 2008. Questioning the empirical basis of the environmental Kuznets curve for CO<sub>2</sub>: New evidence from a panel stationarity test robust to multiple breaks and cross-dependence. Ecological Economics 64 (3), 559 – 574.  
URL <http://www.sciencedirect.com/science/article/pii/S0921800907002042>
- Saxe, A. M., McClelland, J. L., Ganguli, S., 2014. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks.
- Schmalensee, R., Stoker, T., Judson, R., 02 1998. World Carbon Dioxide Emissions: 1950-2050. The Review of Economics and Statistics 80, 15–27.

- Siegelmann, H., Sontag, E., 1995. On the Computational Power of Neural Nets. *Journal of Computer and System Sciences* 50 (1), 132–150.  
 URL <https://www.sciencedirect.com/science/article/pii/S0022000085710136>
- Siegelmann, H. T., 1995. Computation Beyond the Turing Limit. *Science* 268 (5210), 545–548.  
 URL <https://science.sciencemag.org/content/268/5210/545>
- Siegelmann, H. T., Sontag, E. D., 1991. Turing computability with neural nets. *Applied Mathematics Letters* 4 (6), 77–80.  
 URL <https://www.sciencedirect.com/science/article/pii/089396599190080F>
- Stern, D. I., 2004. The Rise and Fall of the Environmental Kuznets Curve. *World Development* 32 (8), 1419 – 1439.  
 URL <http://www.sciencedirect.com/science/article/pii/S0305750X04000798>
- Stern, D. I., Apr 2017. The environmental Kuznets curve after 25 years. *Journal of Bioeconomics* 19 (1), 7–28.  
 URL <https://doi.org/10.1007/s10818-017-9243-1>
- Tsurumi, T., Managi, S., 2015. Environmental Kuznets curve: Economic growth and emission reduction. Cited By 3.  
 URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84943553413&doi=10.4324%2f9781315745800&partnerID=40&md5=65a4dd7392df658a6e0cc2fa45677c83>
- United Nations, 2015. Paris Agreement. Adopted on 12 December 2015 and entered into force on 4 November 2016.
- van Vuuren et al., 2011. The representative concentration pathways: an overview. *Climatic Change* 109 (1), 5.  
 URL <https://doi.org/10.1007/s10584-011-0148-z>
- Vollebergh, H. R., Melenberg, B., Dijkgraaf, E., 2009. Identifying reduced-form relations with panel data: The case of pollution and income. *Journal of Environmental Economics and Management* 58 (1), 27 – 42.  
 URL <http://www.sciencedirect.com/science/article/pii/S0095069609000187>
- Wagner, M., 2008. The carbon Kuznets curve: A cloudy picture emitted by bad econometrics? *Resource and Energy Economics* 30 (3), 388 – 408.  
 URL <http://www.sciencedirect.com/science/article/pii/S0928765507000504>
- Wagner, M., 2015. The Environmental Kuznets Curve, Cointegration and Nonlinearity. *Journal of Applied Econometrics* 30 (6), 948–967.  
 URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2421>
- World Bank, 1992. World population projections 1992 93. Washington.

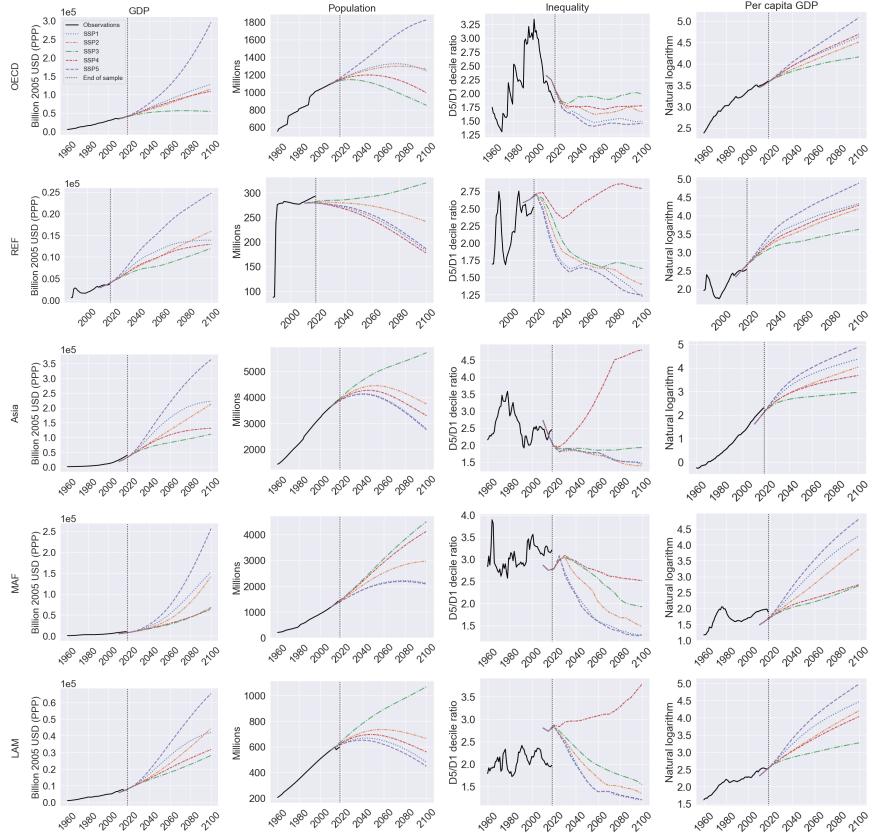
Zhao, X., Du, D., 2015. Forecasting carbon dioxide emissions. *Journal of Environmental Management* 160, 39 – 44.

URL <http://www.sciencedirect.com/science/article/pii/S0301479715300931>

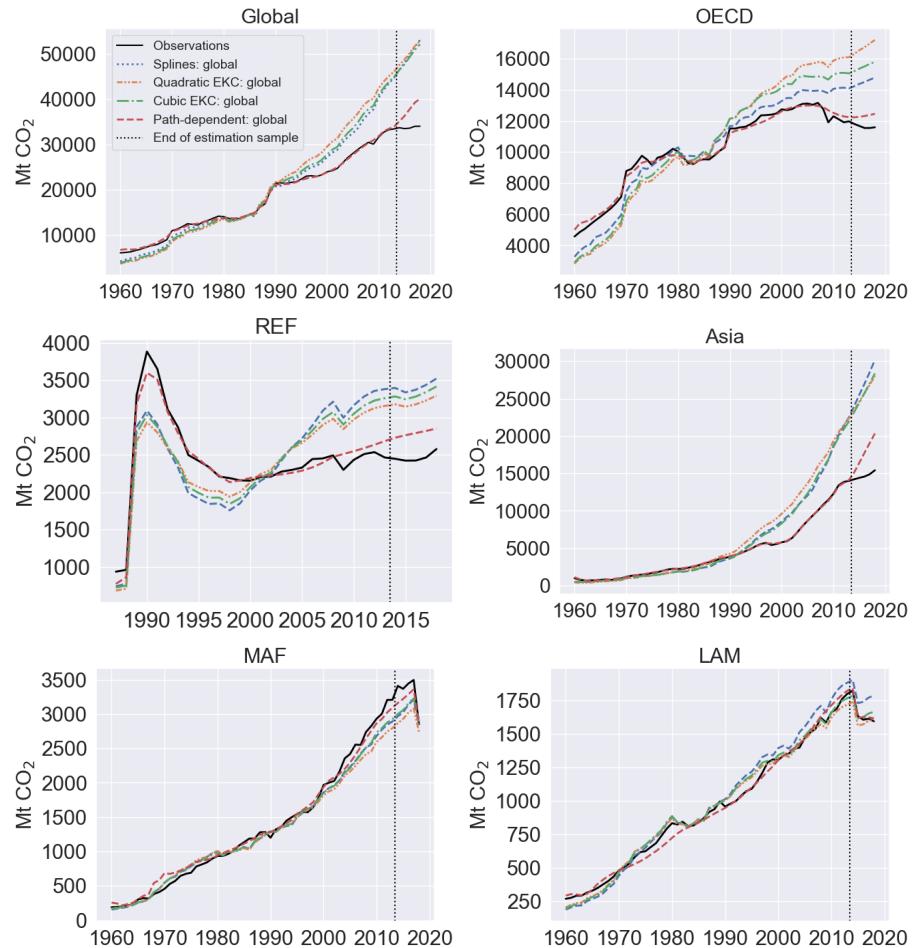
## Appendix

### B.1 Figures

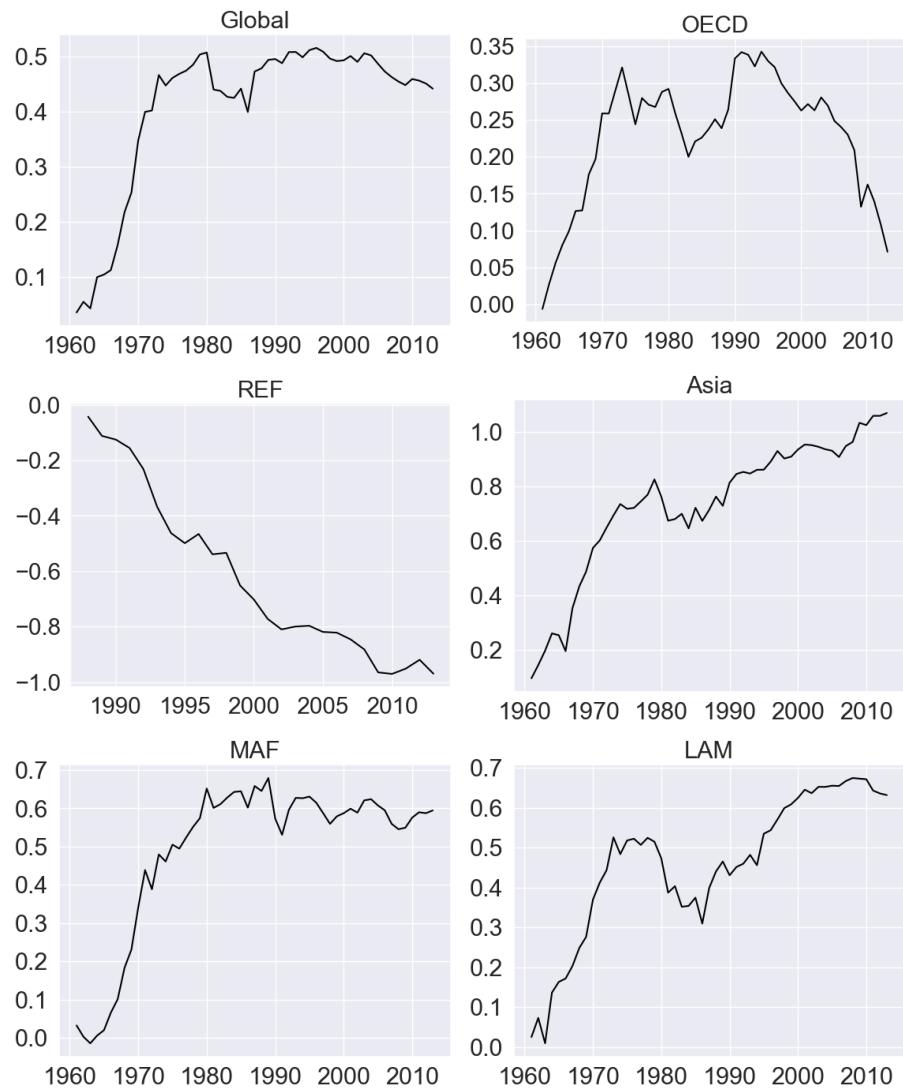
**Figure B.1:** Shared Socioeconomic Pathways basic elements: region level



Note: In the third column, the expression D5/D1 is the ratio of median income to the upper bound value of the first decile. Income is measured by per capita GDP.

**Figure B.2:** Out-of-sample predictions: global model formulations

**Figure B.3:** Estimated time fixed effects from the quadratic EKC model for the estimation sample from 1960 to 2013



Note: Time fixed effects in the top left panel is from a global formulation of the quadratic EKC model. Remaining panels are from region-wise estimation of the quadratic EKC model.

## B.2 Tables

**Table B.1:** Macro-region definitions

<b>OECD</b>		
Albania	Australia	Austria
Belgium	Bosnia and Herzegovina	Bulgaria
Canada	Croatia	Cyprus
Czech Republic	Denmark	Estonia
Finland	France	Germany
Greece	Hungary	Iceland
Ireland	Italy	Japan
Latvia	Lithuania	Luxembourg
Malta	Montenegro	Netherlands
New Zealand	Norway	Poland
Portugal	Puerto Rico	Romania
<b>Serbia</b>	Slovakia	Slovenia
Spain	Sweden	Switzerland
North Macedonia	Turkey	United Kingdom
United States		
<b>REF</b>		
Armenia	Azerbaijan	Belarus
Georgia	Kazakhstan	Kyrgyzstan
Moldova	Russia	Tajikistan
Turkmenistan	Ukraine	Uzbekistan
<b>Kosovo</b>		
<b>Asia</b>		
Afghanistan	Bangladesh	Bhutan
Brunei Darussalam	Cambodia	China
South Korea	Fiji	French Polynesia
India	Indonesia	Laos
Malaysia	Maldives	Micronesia
Mongolia	Myanmar	Nepal
<b>New Caledonia</b>	Pakistan	Papua New Guinea
Philippines	Samoa	Singapore
Solomon Islands	Sri Lanka	Taiwan
Thailand	Timor-Leste	Vanuatu
Vietnam	Tuvalu	Macao
<b>Marshall Islands</b>	Palau	Hong Kong
Tonga	Kiribati	
<b>MAF</b>		
Algeria	Angola	Bahrain
Benin	Botswana	Burkina Faso

Burundi	Cameroon	Cabo Verde
Central African Republic	Chad	Comoros
Congo (DRC)	Congo (RDC)	Côte d'Ivoire
<b>Djibouti</b>	Egypt	Equatorial Guinea
Eritrea	Ethiopia	Gabon
Gambia	Ghana	Guinea
Guinea-Bissau	Iran	Iraq
Israel	Jordan	Kenya
Kuwait	Lebanon	Lesotho
Liberia	Libya	<b>Madagascar</b>
Malawi	Mali	Mauritania
Mauritius	Morocco	Mozambique
Namibia	Niger	Nigeria
Palestine	Oman	Qatar
Rwanda	Saudi Arabia	Senegal
Sierra Leone	<b>Somalia</b>	South Africa
Sudan	Eswatini	<b>Syria</b>
Togo	Tunisia	Uganda
United Arab Emirates	Tanzania	Yemen
Zambia	Zimbabwe	Seychelles
Sao Tome and Principe		

**LAM**

Argentina	Aruba	Bahamas
Barbados	Belize	Bolivia
Brazil	Chile	Colombia
Costa Rica	<b>Cuba</b>	Dominican Republic
Ecudor	El Salvador	<b>Grenada</b>
Guatemala	Guyana	Haiti
Honduras	Jamaica	Mexico
Nicaragua	Panama	Paraguay
Peru	Suriname	Trinidad and Tobago
Uruguay	Venezuela	Saint Vincent
<b>Curaçao</b>	Saint Lucia	<b>Antigua and Barbuda</b>
<b>Dominica</b>	Bermuda	

Note: Region definitions are from the Shared Socioeconomic Pathways (SSPs; Riahi et al., 2017): “OECD” is OECD90 and EU member states and candidates; “REF” is reforming economies Eastern Europe and the former Soviet Union; “Asia” is Asian countries excluding the Middle East, Japan and former Soviet Union states; “MAF” is the Middle East and Africa; and “LAM” is Latin America and the Caribbean. Countries in red are in the SSP database but not in our data set. Countries in green are in our data set but not in the SSP database.

**Table B.2:** Descriptive statistics: mean

	Mean						
	1960	1970	1980	1990	2000	2010	2017
<b>Global</b>							
GDP	3.70	5.49	6.78	7.69	9.01	11.46	13.50
CO <sub>2</sub>	2.51	3.51	3.68	4.19	4.01	4.66	4.64
<b>OECD</b>							
GDP	10.84	16.09	20.47	24.76	29.45	32.62	35.99
CO <sub>2</sub>	8.22	12.08	12.38	12.24	12.31	11.22	10.18
<b>REF</b>							
GDP	NA	NA	NA	10.29	6.52	11.16	12.30
CO <sub>2</sub>	NA	NA	NA	14.06	7.69	8.62	8.44
<b>Asia</b>							
GDP	0.78	1.00	1.37	2.14	3.52	6.54	9.51
CO <sub>2</sub>	0.68	0.65	1.00	1.39	1.79	3.37	3.82
<b>MAF</b>							
GDP	3.23	5.61	6.84	5.06	5.55	6.86	7.21
CO <sub>2</sub>	0.89	1.44	2.09	1.69	2.11	2.46	2.46
<b>LAM</b>							
GDP	5.03	6.54	9.13	8.53	9.84	11.86	12.44
CO <sub>2</sub>	1.32	1.76	2.40	2.24	2.58	2.88	2.73

Note: For given region  $r$  and time period  $t$ , mean values are calculated as the per capita values for that region: Mean  $\equiv \sum_{i \in I_r} \frac{x_{it}}{POP_{it}} \frac{POP_{it}}{POP_t^r} = \sum_{i \in I_r} \frac{x_{it}}{POP_t^r}$ , where  $I_r \subseteq \{1, 2, \dots, N\}$  is the set of indices of countries belonging to region  $r$ ,  $N$  is the total number of countries, POP is population size,  $POP_t^r \equiv \sum_{i \in I_r} POP_{it}$ , and  $x \in \{\text{GDP}, \text{CO}_2\}$ .

**Table B.3:** Descriptive statistics: standard deviation

	Standard deviation						
	1960	1970	1980	1990	2000	2010	2017
<b>Global</b>							
GDP	4.71	7.24	9.43	9.79	11.35	11.67	12.39
CO <sub>2</sub>	4.37	5.81	5.58	5.65	5.20	4.75	4.42
<b>OECD</b>							
GDP	4.16	4.09	5.33	7.88	10.21	10.03	10.51
CO <sub>2</sub>	6.13	6.60	6.22	5.79	6.23	5.29	4.63
<b>REF</b>							
GDP	NA	NA	NA	3.26	2.62	4.62	5.14
CO <sub>2</sub>	NA	NA	NA	4.36	2.93	3.86	4.30
<b>Asia</b>							
GDP	0.51	0.78	1.74	2.28	3.25	4.78	6.11
CO <sub>2</sub>	0.46	0.45	0.82	1.08	1.41	2.69	2.85
<b>MAF</b>							
GDP	2.23	9.60	14.06	7.95	9.16	9.77	10.25
CO <sub>2</sub>	1.55	2.55	3.63	2.98	4.08	4.48	4.53
<b>LAM</b>							
GDP	2.46	2.89	2.90	2.55	2.99	3.36	3.27
CO <sub>2</sub>	1.33	1.44	1.82	1.56	1.75	2.21	1.85

Note: For given region  $r$  and time period  $t$ , standard deviations are calculated using the following relation: Standard deviation  $\equiv \sqrt{\sum_{i \in I_r} \left( \frac{x_{it}}{POP_{it}} - \text{Mean} \right)^2 \frac{POP_{it}}{POP_t^r}}$ , where Mean is from Table B.2,  $I_r \subseteq \{1, 2, \dots, N\}$  is the set of indices of countries belonging to region  $r$ ,  $N$  is the total number of countries, POP is population size,  $POP_t^r \equiv \sum_{i \in I_r} POP_{it}$ , and  $x \in \{\text{GDP}, \text{CO}_2\}$ .

**Table B.4:** Neural network architectures

Neural network architecture	# parameters in path-dependent model		
	Only hidden layers	Global specification ( $R = 1$ )	Regional specification ( $R = 5$ )
(2)	32	34	42
(4)	96	100	116
(8)	320	328	360
(16)	1152	1168	1232
(32)	4352	4384	4512
(2,2)	72	74	82
(4,2)	152	154	162
(4,4)	240	244	260
(8,2)	408	410	418
(8,4)	528	532	548
(8,8)	864	872	904
(16,2)	1304	1306	1314
(16,4)	1488	1492	1508
(16,8)	1952	1960	1992
(16,16)	3264	3280	3344
(32,2)	4632	4634	4642
(32,4)	4944	4948	4964
(32,8)	5664	5672	5704
(32,16)	7488	7504	7568
(32,32)	12672	12704	12832
(2,2,2)	112	114	122
(4,2,2)	192	194	202
(4,4,2)	296	298	306
(4,4,4)	384	388	404
(8,2,2)	448	450	458
(8,4,2)	584	586	594
(8,4,4)	672	676	692
(8,8,2)	952	954	962
(8,8,4)	1072	1076	1092
(8,8,8)	1408	1416	1448
(16,2,2)	1344	1346	1354
(16,4,2)	1544	1546	1554
(16,4,4)	1632	1636	1652
(16,8,2)	2040	2042	2050
(16,8,4)	2160	2164	2180
(16,8,8)	2496	2504	2536
(16,16,2)	3416	3418	3426
(16,16,4)	3600	3604	3620

(16,16,8)	4064	4072	4104
(16,16,16)	5376	5392	5456
(32,2,2)	4672	4674	4682
(32,4,2)	5000	5002	5010
(32,4,4)	5088	5092	5108
(32,8,2)	5752	5754	5762
(32,8,4)	5872	5876	5892
(32,8,8)	6208	6216	6248
(32,16,2)	7640	7642	7650
(32,16,4)	7824	7828	7844
(32,16,8)	8288	8296	8328
(32,16,16)	9600	9616	9680
(32,32,2)	12952	12954	12962
(32,32,4)	13264	13268	13284
(32,32,8)	13984	13992	14024
(32,32,16)	15808	15824	15888
(32,32,32)	20992	21024	21152

Note: " $(a,b,c)$ " indicates a neural network architecture with three hidden layers containing  $a$  units in the first layer,  $b$  in the second, and  $c$  in the third; "# parameters" is the number of free model parameters excluding fixed effects.

**Table B.5:** Adam algorithm

---

**Require:** Step size  $\epsilon$  (0.001)  
**Require:** Exponential decay rates  $\rho_1, \rho_2 \in [0, 1]$  for moment estimates (0.9, 0.999)  
**Require:** Small constant  $\delta$  used for numerical stabilization ( $10^{-8}$ )  
**Require:** Initial parameters  $\theta$

Initialize first moment vector:  $m = 0$   
 Initialize second moment vector:  $v = 0$   
 Initialize time step:  $t = 0$

**while** stopping criterion not met **do**

- $t \leftarrow t + 1$
- Compute gradient:  $g \leftarrow \nabla_{\theta} J(\theta)$
- Update biased first moment estimate:  $m \leftarrow \rho_1 m + (1 - \rho_1)g$
- Update biased second moment estimate:  $v \leftarrow \rho_2 v + (1 - \rho_2)g \odot g$
- Correct bias in first moment:  $\hat{m} \leftarrow m / (1 - \rho_1^t)$
- Correct bias in second moment:  $\hat{v} \leftarrow v / (1 - \rho_2^t)$
- compute update:  $\Delta\theta = -\epsilon \cdot \hat{m} / (\sqrt{\hat{v}} + \delta)$
- Apply update:  $\theta \leftarrow \theta + \Delta\theta$

**end while**

---

Note: Numbers in parentheses are suggested defaults.

The Adam algorithm individually adapts the learning rate of all parameters in two ways. First, its scales them inversely proportional to the square root of an exponentially decaying average of past squared values of the gradient (second moment estimate). In this way, the learning rate of parameters with small partial derivatives of the loss function decreases less rapidly than that of parameters with large partial derivatives. This implies greater progress in more gently sloped regions of the parameter space. Second, to speed up optimization, especially in face of pathological curvature, Adam incorporates momentum by scaling the individual learning rates proportionally to an exponentially decaying average of past values of the gradient (first moment estimate). Finally, Adam includes bias corrections of the moment estimates to account for initialization at the origin.



# 3

C H A P T E R

## NOWCASTING U.S. CO<sub>2</sub> EMISSIONS USING MACHINE LEARNING

**Sebastian Jensen**  
*Aarhus University and CREATEs*

### Abstract

In this paper, I investigate the use of machine learning methods for nowcasting the yearly growth rate of United States carbon dioxide emissions over the period 2000-2019, using a high-dimensional panel of macroeconomic variables sampled at mixed frequencies. To handle the problem of mixed frequencies, I propose to use the frequency alignment transformation from the mixed data sampling regression (MIDAS) literature to represent the data at a common frequency by transforming high-frequency variables into low-frequency vectors through skip-sampling. I find that neural networks, and to a lesser extent tree-based machine learning methods (random forest, bagging, and gradient boosting), are able to utilize the stream of macroeconomic information that becomes available, progressing through the target year, to produce repeatedly more accurate nowcasts of U.S. CO<sub>2</sub> emissions that generally outperform forecasts from univariate time series models and nowcasts from mixed data sampling regression (MIDAS) models.

### 3.1 Introduction

As reported by the Intergovernmental Panel on Climate Change (IPCC) of the United Nations, Carbon dioxide (CO<sub>2</sub>) is the most important greenhouse gas and the key driver of climate change [IPCC] 2014]. Since the pre-industrial era, anthropogenic activity has driven atmospheric concentrations of CO<sub>2</sub> to reach levels unprecedented at least in the last 800,000 years, and effects have been detected throughout the climate system [IPCC] 2014]. To operationalize international long-term obligations to reduce the effects of climate change, such as those put forward by the Paris Agreement<sup>1</sup> under the United Nations Framework Convention on Climate Change (UNFCCC), countries often set yearly targets for CO<sub>2</sub> emissions reductions. However, official numbers on CO<sub>2</sub> emissions are reported with a substantial lag. At the time of writing, August 2021, official estimates of United States CO<sub>2</sub> emissions are available from the inventories of the UNFCCC only until 2019<sup>2</sup>. Nowcasting CO<sub>2</sub> emissions is therefore an important task that can be used to help countries receive more timely feedback on the actions they have taken to reduce their emissions, and thus be more quick to adjust. According to the IPCC, the increase in CO<sub>2</sub> emissions observed over past decades has been driven largely by economic and population growth [IPCC] 2014], highlighting the importance of models that link macroeconomic activity to CO<sub>2</sub> emissions.

I use machine learning methods to nowcast the yearly growth rate of U.S. CO<sub>2</sub> emissions over the period from 2000 to 2019, using a high-dimensional panel of macroeconomic variables sampled at mixed frequencies. As a special case, I also consider one-year ahead forecasting of U.S. CO<sub>2</sub> emissions. [Bennedsen, Hillebrand, and Koopman] 2021c] also consider forecasting and nowcasting of U.S. CO<sub>2</sub> emissions, however, their findings are not directly comparable to those of this paper, as they abstract from problems pertaining to mixed frequencies. To handle the problem of mixed frequencies, I propose to represent the data at a common frequency by using the so-called *frequency alignment* transformation from the mixed data sampling regression (MIDAS) literature, whereby which high-frequency variables are transformed into low-frequency vectors through skip-sampling [Ghysels et al.] 2016]. Although frequency alignment is routinely applied within the MIDAS literature, at least under the hood of popular software implementations [Ghysels et al.] 2016] [Ghysels] 2016], one of the novel insights of this paper is that frequency alignment can be used in other settings too. In particular, popular machine learning methods can be readily applied to mixed frequency data once it has been represented at a common frequency using frequency alignment. However, by using frequency alignment, the number of distinct input signals quickly escalates out of control, and we must solve the curse

---

<sup>1</sup>The goal of the Paris Agreement is to limit the increase in global average temperature to well below 2 degrees Celsius above pre-industrial levels and to pursue efforts to limit the temperature increase to 1.5 degrees Celsius above pre-industrial levels [United Nations] 2015].

<sup>2</sup>Accessible at [di.unfccc.int/time\\_series](http://di.unfccc.int/time_series) last accessed on August 24, 2021.

of dimensionality. In the nowcasting application of this paper, the total number of distinct input signals (leads and lag of the explanatory variables) used in the machine learning methods ranges from 2,053 input signals available for one-year ahead forecasting to 4,105 input signals available for nowcasting by year's end, and only 40 observations are available for estimation and validation in total. The prediction problem considered in this paper, therefore, seems particularly challenging. To solve this prediction problem, I consider the use of tree-based machine learning methods (random forest, bagging, and gradient-boosting) and neural networks, which rely on built-in variable selection mechanisms and feature extraction to solve the curse of dimensionality and are able to account for nonlinear effects of the explanatory variables on emissions and interaction effects between the explanatory variables.

I find that machine learning methods are able to produce repeatedly more accurate nowcasts of U.S. CO<sub>2</sub> emissions, progressing though the target year, which generally outperform forecasts from univariate time series models and nowcasts from MIDAS models. Neural networks tend to produce the most accurate nowcasts of all methods considered, and shallow neural networks with only one or two hidden layers are preferred to deep neural networks with more hidden layers. Focusing on the particular case of a one-hidden-layer neural network and nowcasting by year's end, for illustration, I find the neural network model relies on features derived from most variables in the data set to construct its predictions.

The paper is structured as follows. Section 3.2 discusses how this paper fits into the literature on modeling the relationship between macroeconomic activity and CO<sub>2</sub> emissions and the nowcasting literature more generally. Section 3.3 formally defines the nowcasting problem, emphasizing distinct features of the information set used for nowcasting that makes it a particularly challenging prediction problem. Section 3.4 presents the methods used to nowcast U.S. CO<sub>2</sub> emissions in the experiment of Section 3.5, focusing on tree-based machine learning methods and neural networks. Section 3.6 concludes.

## 3.2 Related literature

The literature on modeling the relationship between macroeconomic activity and carbon dioxide (CO<sub>2</sub>) emissions discusses various "effects" through which macroeconomic activity may affect CO<sub>2</sub> emissions: the *scale effect*, which suggests that economic growth increases emissions, ceteris paribus, through increased use of natural resources and added waste; the *composition effect*, which affects emissions through changes in the output mix; and the *technique effect*, which affects emissions through changes in the state of technology and the input mix (Bennedsen et al. 2021c; Dinda 2004; Stern 2017).

Within climate science, the most common approach to modeling the relationship between macroeconomic activity and CO<sub>2</sub> emissions is to use so-called integrated

assessment models (IAMs), which are large-scale structural models that link key features of the macro economy with key features of the biosphere and the atmosphere [Bouwman et al. 2006] [Messner and Strubegger 1995] [Fujimori et al. 2012] [Calvin 2011] [Luderer et al. 2013] [Emmerling et al. 2016]. However, IAMs are computationally intensive, rely on a large number of structural and parametric assumptions, and parameter values are typically set using a mixture of judgment and calibration. Therefore, less computationally demanding, reduced-form models with few parametric restrictions and a well-developed estimation theory have also been proposed [Auffhammer and Steinhauser 2012] [Wagner 2008] [2015] [Bennedsen et al. 2021c] and are often concerned with a possible turning point in the relationship between economic growth and emissions, referred to as the environmental Kuznets curve [Grossman and Krueger 1991] [Arrow et al. 1995]. Both the IAMs and the reduced-form models typically rely on a highly aggregated measure of economic activity, such as gross domestic product (GDP), and, typically, focus on projecting CO<sub>2</sub> emissions several decades into the future [Holtz-Eakin and Selden 1995] [Schmalensee et al. 1998] [Zhao and Du 2015] [Bennedsen et al. 2021b]. By stark contrast, this paper focuses on nowcasting CO<sub>2</sub> emissions using machine learning methods and a high-dimensional panel of macroeconomic variables.

The term *nowcasting* originated in meteorology and refers to prediction of the very near future, the present, and the very recent past. Many important variables used to describe society and the macro economy, such as CO<sub>2</sub> emissions, GDP, and other macroeconomic variables, are sampled at low frequencies, typically on a quarterly basis or on a yearly basis and released with a substantial delay. Nowcasting exploits the informational content in a set of explanatory variables that are sampled more frequently than the variable of interest to update an initial forecast before its realized value is released. Nowcasting is, therefore, intrinsically a mixed-frequency prediction problem. As discussed by [Banbura, Giannone, and Reichlin 2010], it is key to use the most up-to-date high-frequency information, often released in a non-synchronous manner and with varying publication lags, leading to a so-called “ragged” edge with missing values in the end of the sample.

The simplest approach to nowcasting with mixed frequency data is to use so-called bridge equations [Baffigi, Golinelli, and Parigi 2004] [Diron 2008]. Bridge equations are simple, linear regression models that relate (“bridge”) the low-frequency variable of interest to low-frequency aggregates of the high-frequency explanatory variables, constructed in an initial step, often using equal weights. However, much information may be lost when aggregating to lowest frequency. Therefore, the unrestricted mixed data sampling regression (U-MIDAS) model suggests to use all leads and lags of the high-frequency explanatory variables, available at the time of prediction, in a linear regression model [Foroni, Marcellino, and Schumacher 2015]. Whereas this approach utilizes all available information, it quickly becomes infeasible to estimate the U-MIDAS model with ordinary least squares (OLS) methods in

situations with a high-dimensional set of explanatory variables or large differences in sampling frequency between the variable of interest and the explanatory variables, due to parameter proliferation. To alleviate parameter proliferation issues, the U-MIDAS model can be estimated using regularized linear regression methods. This paper suggests to use the least absolute shrinkage and selection operator (LASSO; Tibshirani [1996]) to estimate the U-MIDAS model. The restricted MIDAS model (henceforth MIDAS model) seeks to alleviate parameter proliferation issues by restricting the parameters of the U-MIDAS model, using parsimoniously parametrized distributed lag polynomials (Ghysels et al. [2004] [2005] [2007] Andreou et al. [2010]). Note the MIDAS model can be understood as a bridge equation where low-frequency aggregates of the explanatory variables are constructed using data-driven aggregation schemes (Ghysels et al. [2016]). The price to be paid for introducing parametrized lag polynomials in the MIDAS model is that the model becomes nonlinear in its parameters, and nonlinear least squares (NLS) estimation is required. Using a high-dimensional set of explanatory variables, it may not be feasible to estimate the MIDAS model with NLS methods even after parsimoniously parametrizing the distributed lag polynomials, and initial dimensionality reduction may be required. This paper suggests to select relevant variables for the MIDAS model in an initial step using the *targeting* approach of Bai and Ng [2008]. An alternative approach would be to extract common factors from the set of explanatory variables, then use the common factors as input variables in the MIDAS model (Marcellino and Schumacher [2010]).

The aforementioned models are single-equation models where parameters depend on the forecast horizon. Since nowcasts are typically constructed repeatedly as new information arrives, a separate version of the models needs to be estimated for each forecast horizon. An arguably more sophisticated approach is to specify an entire system of equations, then cast it in state-space form and use the Kalman filter to impute missing values and provide linear projections of the variable of interest, as suggested by Giannone, Reichlin, and Small [2008]; Banbura et al. [2010]; Doz, Giannone, and Reichlin [2011]; Bańbura and Modugno [2014]. To allow for a high-dimensional set of explanatory variables, the general framework is typically a factor model along the lines of Forni, Hallin, Lippi, and Reichlin [2000]; Stock and Watson [2002]. Since these state-space models summarize the joint dynamics of the variable of interest and the explanatory variables, they can be readily used across forecast horizons, and to interpret nowcast revisions in terms of news embedded in consecutive data releases. Although the Kalman filter provides optimal linear projections under correct specification, these state-space models are prone to misspecification, as a full system of equations needs to be specified. Bai, Ghysels, and Wright [2013] also show that, in some cases, the MIDAS model even provides an exact representation of the Kalman filter; and in cases where the MIDAS model does not provide an exact approximation of the Kalman filter, approximation errors are typically small.

This paper proposes a machine learning approach to nowcasting using all leads

and lags of the high-frequency explanatory variables, available at the time of prediction, in machine learning methods that themselves must decide how to best utilize the information. I propose to use the frequency alignment transformation from the MIDAS literature to represent the data at a common frequency. The appeal of the machine learning approach is that all available information can then be fed into the machine learning methods without need of initial dimensionality reduction or low-frequency aggregation, and it avoids pre-imposing a lot of parametric structure. The machine learning methods solve the curse of dimensionality through built-in variable selection mechanisms and feature extraction, and in contrast to the aforementioned approaches, are also able to account for nonlinear effects of the explanatory variables on the variable of interest and interaction effects between the explanatory variables.

### 3.3 The nowcasting problem

Let  $y_t \in \mathbb{R}$  be a variable of interest at time  $t$ , and assume that  $y_t$  is sampled at a fixed sampling frequency (the interval of reference). Let  $x_{t_q}^q \in \mathbb{R}^{n_q}$  be a vector of explanatory variables sampled  $q$  times faster, and let  $x_{t_m}^m \in \mathbb{R}^{n_m}$  be a vector of explanatory variables sampled  $m$  times faster. To fix ideas, assume  $y$  is sampled yearly, and that  $x^q$  and  $x^m$  denote quarterly and monthly variables, respectively. These definitions are consistent with the data used in the empirical application of Section 3.5. This paper defines forecasting and nowcasting of  $y_t$  as the problem of finding the regression function

$$g : \Omega_v \rightarrow \mathbb{R}, \text{ which characterizes the conditional expectation } \mathbb{E}[y_t | \Omega_v], \quad (3.1)$$

where  $\Omega_v$  denotes the information set available at time  $v$ ,  $v \in \mathbb{Q}^+$ . By having  $v$  a rational number, equation (3.1) allows predictions to be made *within* a year. This paper defines the prediction problem in (3.1) to be one of forecasting if  $v \leq t - 1$ , and one of nowcasting if  $v > t - 1$ . Some distinct features of  $\Omega_v$  are encountered in practice: mixed frequencies, a "ragged edge", and high dimensionality. In the current setting,  $\Omega_v$  can be represented as

$$\Omega_v = \Omega_v^m \cup \Omega_v^q \cup \Omega_v^y,$$

where

$$\begin{aligned} \Omega_v^m &= \left\{ x_{it_m}^m, i = 1, \dots, n_m, t_m = \frac{1}{m}, \frac{2}{m}, \dots, \frac{T_{iv}}{m} \right\}, \\ \Omega_v^q &= \left\{ x_{jt_q}^q, j = 1, \dots, n_q, t_q = \frac{1}{q}, \frac{2}{q}, \dots, \frac{T_{jv}}{q} \right\}, \\ \Omega_v^y &= \{y_t, t = 1, 2, \dots, T_v\}, \end{aligned}$$

and  $\Omega_v^m$  and  $\Omega_v^q$  denote the information set available from the monthly and the quarterly variables, respectively, and  $\Omega_v^y$  denotes the information set available from the

history of the dependent variable itself. The total number of observations on the  $i$ -th monthly variable and the  $j$ -th quarterly variable in vintage  $v$  is denoted by  $T_{iv}^m$  and  $T_{jv}^q$ , respectively, and  $T_v$  denotes the total number of observations on the dependent variable. Since data is released in a non-synchronous manner with different degrees of delay, some variables will have missing values in the end of the sample, leading to a ragged edge. The total number of monthly and quarterly variables is denoted by  $n_m$  and  $n_q$ , respectively. It is not always clear, a priori, what other variables are important for predicting the variable of interest, and with large data sets often readily available, it is common practice to include a high-dimensional set of explanatory variables sampled at mixed frequencies in  $\Omega_v$ , of which not all the explanatory variables, typically, contain any predictability. The empirical application of Section 3.5 is characterized by  $n_m + n_q \gg T_v$ . To alleviate the curse of dimensionality, the prediction models used for  $g$  in 3.1 should entail a form of dimensionality reduction through either variable selection, feature extraction, or both. To handle the problem of mixed frequencies, I propose to use the frequency alignment transformation from the MIDAS literature. The next section discusses how machine learning methods can be used to solve the curse of dimensionality, and how they can be readily applied to mixed frequency data after frequency alignment.

As discussed in Banbura et al. (2010), practitioners typically perform a series of nowcasts that are repeatedly updated as new information arrives:  $\mathbb{E}[y_t | \Omega_{v_1}]$ ,  $\mathbb{E}[y_t | \Omega_{v_2}]$ , ..., where  $v_1 < v_2, \dots$ , and  $v_1, v_2, \dots$  refer to dates of consecutive data releases. In practice, the time between data releases is often short and may change over time, implying the  $v$ -s may be both high-frequency and irregularly spaced. The first nowcast is typically made using only little additional information relative to the preceding forecast. However, as new data is released and additional information on the target period accrues, nowcasts tend to get increasingly precise. The relationship between the information sets  $\Omega_{v_1}$  and  $\Omega_{v_2}$ ,  $v_1 < v_2$ , can be illustrated as

$$\Omega_{v_1} \subseteq \Omega_{v_2} \text{ and } \Omega_{v_2} \setminus \Omega_{v_1} = \left\{ x_{iT_{iv_2}}^m, i \in \mathbb{I}_{v_2}; x_{jT_{jv_2}}^q, j \in \mathbb{J}_{v_2} \right\}.$$

Note the information set expands at time  $v_2$  as new observations on a set of variables are released.<sup>3</sup> In practice, past data might also get revised, and one can even imagine having  $\Omega_{v_2} \subseteq \Omega_{v_1}$  if some series are discontinued or removed.

### 3.3.1 Missing values and outliers

As already discussed,  $\Omega_v$  may contain missing values in the end of the sample due to a ragged edge. The information set  $\Omega_v$  may also contain missing values in the beginning of the sample, as some series may not start until later in the sample period.

---

<sup>3</sup>The dependent variable could be included in a data release. I abstract from this case not to complicate notation.

In addition, I classify observations that deviate from the sample median by more than ten interquartile ranges as outliers, following McCracken and Ng (2016, 2020), and also treat these outliers as missing values. In the empirical application of Section 3.5, all missing values are estimated using the approach of McCracken and Ng (2016, 2020). Below, I illustrate the approach for the monthly variables in  $\Omega_v^m$ . The same approach is used for the quarterly variables in  $\Omega_v^q$ .

Let  $T_v^m$  denote the number of months in  $\Omega_v^m$ , defined as  $T_v^m \equiv \max\{T_{1v}^m, \dots, T_{n_m v}^m\}$ . To estimate the missing values in  $\Omega_v^m$ , it is assumed that all monthly variables admit a factor model representation:

$$x_{it_m}^m = \lambda_i^{m\top} f_{t_m}^m + e_{it_m}^m, \quad i = 1, \dots, n_m; t_m = \frac{1}{m}, \frac{2}{m}, \dots, \frac{T_v^m}{m}, \quad (3.2)$$

where  $f_{t_m}^m \in \mathbb{R}^r$  is column vector of latent and common factors, and  $\lambda_i^m \in \mathbb{R}^r$  is a column vector of cross-sectional-specific factor loadings. The idiosyncratic disturbances  $e_{it_m}^m$  are allowed to be both serially correlated and weakly cross-sectionally correlated. Equation (3.2) can, therefore, be considered a serially correlated version of the approximate factor model of Chamberlain and Rothschild (1983). The expectation maximization (EM) algorithm is used together with standard principal component analysis (PCA) to estimate missing values. Initially, all monthly variables are studentized to have zero mean and unit variance, and missing values are initialized to zero (the unconditional mean) to obtain a balanced panel. For a given number of factors  $r$ , standard PCA, with normalization  $\lambda^{m\top} \lambda^m = I_r$ , is used to retrieve factor estimates  $\hat{f}_{t_m}^m$ ,  $t_m = \frac{1}{m}, \frac{2}{m}, \dots, \frac{T_v^m}{m}$ , and factor loadings  $\hat{\lambda}_i^m$ ,  $i = 1, \dots, n_m$ . In the E-step, estimates of the missing values are updated using  $\hat{x}_{it_m}^m = \hat{\lambda}_i^{m\top} \hat{f}_{t_m}^m$ . This value is then multiplied by the sample standard deviation, and the sample mean is added back. In the M-step, missing-value estimates from the E-step are treated as observations. The mean and the variance of the compete sample are recalculated, and the data is re-studentized. PCA is then reapplied to obtain new factor estimates and new factor loadings. The procedure continues by iterating the E-step and the M-step until convergence of the factor estimates.

Following McCracken and Ng (2016, 2020), the number of factors  $r$  is selected by using the  $PC_2$  criterion of Bai and Ng (2002). This criterion is constructed as the sum of squared residuals plus a penalty term that is added to retain model parsimony. The  $PC_2$  criterion relies on the penalty term  $\frac{n_m + T_v^m}{T_v^m} \log(\min(n_m, T_v^m))$ , which is shown by Bai and Ng (2002) to have good finite sample properties.

The MATLAB code for estimating missing values using the approach above is readily available through Michael W. McCracken's personal website.<sup>4</sup> Alternative approaches to dealing with the issue of missing values due to a ragged edge have been proposed in the literature. Most notably, Doz et al. (2011) specify a complete and parametric dynamic factor model representation of the variables in  $\Omega_v$ , then

---

<sup>4</sup> Accessible at: [research.stlouisfed.org/econ/mccracken/fred-databases/](https://research.stlouisfed.org/econ/mccracken/fred-databases/) last accessed on April 20, 2021.

use the Kalman filter to estimate missing values. Altissimo, Cristadoro, Forni, Lippi, and Veronese (2010) propose a simple, vertical realignment of the data to obtain a balanced panel. However, this latter approach does not immediately account for missing values in the beginning or in the middle of the sample, and data availability determines cross-correlations between variables.

### 3.4 Methodology

The section starts by introducing a general one-step ahead prediction model, using mixed frequency data. Next, I present a set of standard univariate time series models and discuss the restricted and the unrestricted MIDAS model. These models will serve as benchmarks in the empirical application of the next section. When discussing the MIDAS model, I demonstrate how the frequency alignment transformation can be used to represent mixed frequency data at a common frequency by transforming high-frequency variables into low-frequency vectors. One of the key insights of this paper is that frequency alignment can also be used to allow most machine learning methods to be readily applied to mixed-frequency data. I discuss the use of tree-based machine learning methods and neural networks, which are the main focus of the paper. The Machine learning methods can be considered single equation models much like the U-MIDAS and the MIDAS models, which are therefore considered more natural benchmarks than the state-space models of Giannone et al. (2008); Banbura et al. (2010).

Let  $h$  denote a fraction between 0 and 1, and let  $y_{t+1|t+h} \equiv \mathbb{E}[y_{t+1} | \Omega_{t+h}]$  denote the conditional expectation of  $y_{t+1}$  given information available at time  $t+h$ . The dependent variable and the information set available at time  $t+h$  is defined as in the previous section. The conditional expectation is a one-step ahead forecast if  $h=0$ , and a nowcast if  $h>0$ . Let  $\ell_h^m = \lfloor mh \rfloor$  denote the number of monthly leads available  $h$ -th of a year into year  $t+1$  (target year), and let  $\ell_h^q = \lfloor qh \rfloor$  denote the number of quarterly leads available  $h$ -th of a year into year  $t+1$ . The general one-step ahead prediction model can then be represented as

$$y_{t+1|t+h} = g(x_{t+h}), \quad (3.3)$$

where  $x_{t+h} \equiv (y_{t-}, x_{t-}^q, x_{t-}^m, x_{t+}^q, x_{t+}^m)$  is a high-dimensional vector that collects all

leads and lags of the explanatory variables available at the time of prediction:

$$\begin{aligned} y_{t-} &\equiv \left( y_t, \dots, y_{t-p_y+1} \right), \\ x_{t-}^q &\equiv \left( x_t^q, x_{t-1/q}^q, \dots, x_{t-(p_q-1)/q}^q \right), \\ x_{t-}^m &\equiv \left( x_t^m, x_{t-1/m}^m, \dots, x_{t-(p_m-1)/m}^m \right), \\ x_{t+}^q &\equiv \left( x_{t+\ell_h^q/q}^q, x_{t+(\ell_h^q-1)/q}^q, \dots, x_{t+1/q}^q \right), \\ x_{t+}^m &\equiv \left( x_{t+\ell_h^m/m}^m, x_{t+(\ell_h^m-1)/m}^m, \dots, x_{t+1/m}^m \right), \end{aligned}$$

where  $p_q$  and  $p_m$  denote the number of quarterly and monthly lags, respectively, available at the time of prediction, and  $p_y$  denotes the number of lags of the dependent variables, available at the time of prediction. In what follows, I restrict attention to lagged values from the previous year only<sup>5</sup> and fix  $p_y \equiv 1$ ,  $p_q \equiv q$ , and  $p_m \equiv m$ . The remainder of this section presents different models that impose different assumption on the form of  $g$ , and present different ways to solve the curse of dimensionality.

### 3.4.1 Univariate time series models

Following [Bennedsen et al. (2021c)], I consider a set of standard univariate time series models:

- A constant growth (Cst) model, where predictions are constructed using the relation  $y_{t+1|t+h} = \frac{1}{t} \sum_{\tau=1}^t y_{\tau}$ .
- A random walk (RW) model, where predictions are constructed using the relation  $y_{t+1|t+h} = y_t$ .
- An autoregressive (AR) model of order one, where predictions are constructed using the relation  $y_{t+1|t+h} = \alpha y_t$ ,  $\alpha \in \mathbb{R}$ .
- A moving average (MA) model of order one, where predictions are constructed using the relation  $y_{t+1|t+h} = \pi \varepsilon_t$ ,  $\pi \in \mathbb{R}$  and  $\varepsilon_t$  is zero-mean white noise.
- An autoregressive and moving average (ARMA) model of order (1, 1), where predictions are constructed using the relation  $y_{t+1|t+h} = \alpha y_t + \pi \varepsilon_t$ ,  $\alpha, \pi \in \mathbb{R}$  and  $\varepsilon_t$  is zero-mean white noise.

Clearly, univariate time series models cannot be used for nowcasting, only forecasting. Nonetheless, univariate time series models constitute simple and important benchmarks. In particular, the constant growth model serves as main benchmark in the nowcasting experiment of Section 3.5.

---

<sup>5</sup>The constant growth model constitutes the only exception to this rule.

### 3.4.2 MIDAS

Assuming that  $g$  in (3.3) is linear, the one-step ahead prediction model then takes the form:

$$y_{t+1|t+h} = \mu + \alpha y_t + \sum_{j=1}^{n_q} \sum_{k=0}^{\ell_h^q + q - 1} \beta_{jk}^q x_{j,t+(\ell_h^q - k)/q}^q + \sum_{i=1}^{n_m} \sum_{k=0}^{\ell_h^m + m - 1} \beta_{ik}^m x_{i,t+(\ell_h^m - k)/m}^m, \quad (3.4)$$

where  $\mu \in \mathbb{R}$  is an overall intercept,  $\alpha \in \mathbb{R}$  is an autoregressive slope coefficient and  $\beta_{jk}^q, \beta_{ik}^m \in \mathbb{R}$  are slope coefficients. The prediction model in (3.4) can be understood as an unrestricted MIDAS (U-MIDAS) model, and can be estimated using OLS provided the number of free parameters does not exceed the number of observations available for estimation. However, in situations where (3.4) contains a large number of explanatory variables, a large number of leads and lags of the explanatory variables, or there is large a difference in sampling frequency between the dependent variable and the explanatory variables, parameter proliferation leaves OLS estimation infeasible. Below, I discuss how the U-MIDAS model can be estimated in such situations by using regularized linear regression methods, and how the restricted MIDAS model alleviates some of the parameter proliferation issues by imposing parametric restrictions on the parameters in (3.4). First, I briefly introduce the concept of frequency alignment.

#### 3.4.2.1 Frequency alignment

It is instructive to demonstrate how (3.4) can be represented as a standard time series regression model where all variables are observed at a common frequency, using matrix notation. The transformation is referred to as frequency alignment [Ghysels et al., 2016]. Suppose we have available the information set  $\Omega_\nu$ , characterized in Section 3.3. Then, equation (3.4) can be written as

$$\begin{bmatrix} y_{2|1+h} \\ \vdots \\ y_{T_\nu|T_\nu-1+h} \end{bmatrix} = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_{T_\nu-1} \end{bmatrix} \alpha + \sum_{j=1}^{n_q} X_j^q \begin{bmatrix} \beta_{j,0}^q \\ \vdots \\ \beta_{j,\ell_h^q + q - 1}^q \end{bmatrix} + \sum_{i=1}^{n_m} X_i^m \begin{bmatrix} \beta_{i,0}^m \\ \vdots \\ \beta_{i,\ell_h^m + m - 1}^m \end{bmatrix}, \quad (3.5)$$

where  $T_\nu$  denotes the total number of yearly observations in  $\Omega_\nu$ , and  $X_j^q$  and  $X_i^m$  are obtained by skip-sampling from  $\Omega_\nu^q$  and  $\Omega_\nu^m$ , respectively:

$$X_j^q = \begin{bmatrix} x_{j,1+\ell_h^q/q}^q & \cdots & x_{j,1/q}^q \\ \vdots & \ddots & \vdots \\ x_{j,T_\nu-1+\ell_h^q/q}^q & \cdots & x_{j,T_\nu-2+1/q}^q \end{bmatrix}, \quad X_i^m = \begin{bmatrix} x_{i,1+\ell_h^m/m}^m & \cdots & x_{i,1/m}^m \\ \vdots & \ddots & \vdots \\ x_{i,T_\nu-1+\ell_h^m/m}^m & \cdots & x_{i,T_\nu-2+1/m}^m \end{bmatrix}. \quad (3.6)$$

Here, I have used the linear model in (3.4) to illustrate frequency alignment. However, it is clear that once all variables are transformed to a common frequency, we can

readily apply most regression methods. In particular, frequency alignment allows us to readily apply the tree-based machine learning methods and neural networks, presented below, to mixed frequency data. Note also that frequency alignment in a sense increases the cross-sectional dimension of the data. In what follows, I present different models that represent different ways to solve the curse of dimensionality.

### 3.4.2.2 Unrestricted MIDAS with regularization

As discussed above, it is often not feasible to estimate the U-MIDAS model in (3.4) by OLS due to parameter proliferation. However, even in a situation where OLS estimation is not feasible, the model can be estimated using regularized linear regression methods. Regularized regression solves the curse of dimensionality by minimizing the objective function under some constraint that bounds the norm of the parameter vector from above.

Assuming that we have the information set  $\Omega_v$ , defined in Section 3.3, available for estimation, the Lagrangian form of the regularized linear regression problem based on the U-MIDAS model in (3.4) can be represented as:

$$J(\beta) = \sum_{t=1}^{T_v} (y_{t+1} - \hat{y}_{t+1|t+h}(x_{t+h}))^2 + R(\beta), \quad (3.7)$$

where  $x_{t+h}$  is defined above,  $\hat{y}_{t+1|t+h}$  is given by (3.4) and depends on the estimated parameters  $\hat{\mu}$  and  $\hat{\beta} \equiv (\hat{\alpha}, \hat{\beta}^q, \hat{\beta}^m)$ , and  $R(\cdot)$  is a penalty term that depends on the norm of the estimated parameter vector  $\hat{\beta}$ . Popular choices for specification of the penalty term are provided by the LASSO estimator [Tibshirani 1996], the Ridge estimator [Hoerl and Kennard 1970], the elastic net estimator [Zou and Hastie 2005], the group LASSO estimator [Yuan and Lin 2006], and the sparse-group LASSO estimator [Simon, Friedman, Hastie, and Tibshirani 2013]. I specify the penalty term based on the LASSO estimator due to its popularity. The LASSO estimator specifies the penalty term in (3.7) as  $R(\beta) = \lambda \|\beta\|_1$ , where  $\|\cdot\|_1$  denotes the  $L^1$  norm. The hyperparameter  $\lambda$  controls the bias-variance tradeoff and determines the degree of sparsity<sup>6</sup>. I tune  $\lambda$  using a validation sample; see Section 3.5.2. Geometrically, the LASSO estimator minimizes the sum of squared errors objective function over a rhomboid  $\{\beta : \|\beta\|_1 \leq C\}$ , and for this reason, LASSO encourages variable selection by setting some parameters in  $\beta$  exactly to zero [Tibshirani 1996]. Below, I argue the variable selection mechanism embedded in LASSO can also be used to select variables for the restricted MIDAS model.

### 3.4.2.3 Restricted MIDAS

The central idea of the restricted MIDAS model (henceforth MIDAS model) is to reduce the number of free parameters in (3.4) by imposing additional structure. The

---

<sup>6</sup>The hyperparameter  $\lambda$  is sometimes said to control the degree of shrinkage.

MIDAS model can be related to the temporal aggregation literature by being represented in terms of low-frequency aggregates, which is similar to the idea underlying bridge equations. The difference is that bridge equations aggregate to lowest frequency using a simple Pythagorean mean function, whereas the MIDAS model relies on data-driven, albeit parametric, and parsimonious aggregation functions. I rely on the following representation of the MIDAS model:

$$y_{t+1|t+h} = \tilde{\mu} + \tilde{\alpha} y_t + \sum_{j=1}^{n_q} \tilde{\beta}_j^q \bar{x}_{j,t+h}^q + \sum_{i=1}^{n_m} \tilde{\beta}_i^m \bar{x}_{i,t+h}^m, \quad (3.8)$$

where  $\bar{x}^q$  and  $\bar{x}^m$  denote data-driven, low-frequency aggregates:

$$\forall j, \bar{x}_{j,t+h}^q = \sum_{s=1}^{\ell_h^q+q} \omega_j^q(\delta_j^q; s) x_{j,t-1+s/q}^q, \quad \forall i, \bar{x}_{i,t+h}^m = \sum_{s=1}^{\ell_h^m+m} \omega_i^m(\delta_i^m; s) x_{i,t-1+s/m}^m. \quad (3.9)$$

The low-frequency aggregates are constructed using a weighting function  $\omega$  that depends on a parameter vector  $\delta$ . The aggregation weights should be non-negative, and for identification of the slope-coefficients in (3.8), they should add up to one. To satisfy these constraints, the weighting functions are defined in the following way:

$$\forall j, s, \omega_j^q(\delta_j^q; s) = \frac{\psi_j^q(\delta_j^q; s)}{\sum_{\tau=1}^{\ell_h^q+q} \psi_j^q(\delta_j^q; \tau)}, \quad \forall i, s, \omega_i^m(\delta_i^m; s) = \frac{\psi_i^m(\delta_i^m; s)}{\sum_{\tau=1}^{\ell_h^m+m} \psi_i^m(\delta_i^m; \tau)}, \quad (3.10)$$

for underlying functions  $\psi$ . Provided the underlying functions  $\psi$  are non-negative (and the denominators in (3.10) are positive), the weighting functions in (3.10) are non-negative. One can be very creative in the design of weighting functions (Ghysels et al. [2016]). Following Ghysels et al. (2005) [2007], I use the Exponential Almon lag polynomial with two free parameters to parametrize all weighting functions:  $\psi(\delta; s) = \exp(\delta_1 s + \delta_2 s^2)$  for  $\delta_1, \delta_2, s \in \mathbb{R}$ . Even with just two free parameters, the Exponential Almon lag polynomial can take on various shapes, including decreasing, increasing, or hump-shaped patterns, and automatically determines the number of leads and lags of the explanatory variables that is effectively included in the model (Ghysels et al. [2005]). Note also that the arithmetic mean, often used in bridge equations, is obtained as a special case for  $\delta_1 = \delta_2 = \dots = \delta_c = 0$ . However, pre-imposing this flat weighting scheme generally leads to omitted variable bias (Andreou et al. [2010]). The price to be paid for use of the Exponential Almon lag polynomial is that the parameters  $\delta$  enter nonlinearly in (3.8). Estimation, therefore, relies on nonlinear least squares (NLS).

Note that I construct low-frequency aggregates jointly from the leads and the lags of the explanatory variables. One can also imagine constructing low-frequency

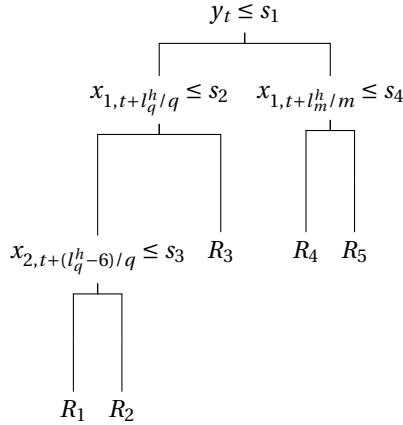
aggregates separately from the leads and the lags of the explanatory variables, respectively. However, NLS estimation of (3.8) requires the number of observations available for estimation to be greater than the number of slope coefficients. In the empirical application of Section 3.5 few observations are available for estimation, which puts a low ceiling on the number of explanatory variables that can be included in the model. If low-frequency aggregates were to be constructed separately from the leads and the lags of the explanatory variables, respectively, this ceiling would be lowered even further. Alluding to earlier, I use LASSO to select variables for the MIDAS model in the empirical application of Section 3.5. For a given number of explanatory variables to select, I increase the regularization of LASSO until only input signals (leads and lags) from the given number of explanatory variables are remaining. All leads and lags of those variables are then included in the MIDAS model. This is similar to the targeting approach of Bai and Ng (2008). I consider selecting 10, 20, and 30 explanatory variables for the MIDAS model.

### 3.4.3 Tree-based methods for regression

Thus far, we have focused on parametric methods that impose a lot of structure, including linearity, on the general prediction model in (3.3). This section introduces tree-based machine learning methods, which constitute a set of fully nonparametric methods that solve the curse of dimensionality through built-in variable selection. They also allow the regression function to be nonlinear and account for interaction effects between the explanatory variables. This section starts by introducing the notion of a regression tree. It then discusses how multiple regression trees can be pooled together into an ensemble for improved predictive performance. Various ways of pooling together regression trees have been proposed in the literature. I focus on the random forest algorithm of Breiman (2001) and the gradient boosting algorithm of Friedman (2001). They appear to be the most popular, tree-based ensemble methods.

The regression tree presents a conceptually simple, yet powerful, method that generalizes the intuition of a decision tree to regression. The regression tree works by partitioning the input space into subspaces, then fitting a constant within each subspace. Different algorithms for partitioning the input space have been proposed in the literature. I rely on the popular classification and regression tree (CART) algorithm of Breiman, Friedman, Stone, and Olshen (1984), which is based on a recursive binary partitioning of the input space into hyperrectangular and disjoint subspaces. Figure 3.1 illustrates a hypothetical regression tree for the prediction model in (3.3).

Consider the general one-step ahead prediction model in (3.3). Assume that we have an available information set  $\Omega_v$ , characterized in Section 3.3. Assume also the input space spanned by  $x_{t+h} = (y_{t-}, x_{t-}^q, x_{t-}^m, x_{t+}^q, x_{t+}^m)$  has been partitioned into  $M$  disjoint and hyperrectangular subspaces  $R_1, R_2, \dots, R_M$ . The regression tree models the dependent variable  $y_{t+1}$  by a constant  $c_m$  in each subspace, and assumes the

**Figure 3.1:** Regression tree

Note: The figure illustrates a hypothetical regression tree for the one-step ahead prediction problem in (3.3). At each junction, observations satisfying the given condition are assigned to the left branch, and remaining observations are assigned to the right branch. In this hypothetical regression tree, the input space is initially split into two subspaces at  $y_t = s_1$ . The subspace defined by  $y_t \leq s_1$  is then split into two subspaces at  $x_{1,t+l_q^h/q} = s_2$ , and the subspace defined by  $y_t > s_1$  is split into two subspaces at  $x_{1,t+l_m^h/m} = s_4$ . The subspace defined by both  $y_t \leq s_1$  and  $x_{1,t+l_q^h/q} \leq s_2$  is split into two subspaces at  $x_{2,t+(l_q^h-6)/q} = s_3$ , after which it is optimal to make no further splits. The final tree has five leaves, corresponding to the subspaces  $R_1, R_2, \dots, R_5$ , and a depth (the length of the longest path from *root* to *leaf*) of three.

function  $g$  in equation (3.3) can be represented as

$$g(x_{t+h}; \Theta) = \sum_{m=1}^M c_m \mathbb{1}_{\{x_{t+h} \in R_m\}}, \quad (3.11)$$

where  $\Theta$  characterizes the regression tree in terms of split-variables, split points, and constants (*leaf values*). Here, the objective function is taken to be squared error loss, and the estimator for the leaf values  $c_m$  is therefore the average of the dependent variable within the subspace:

$$\hat{c}_m = \frac{\sum_{t=1}^{T_v} y_t \mathbb{1}_{\{x_{t+h} \in R_m\}}}{\sum_{t=1}^{T_v} \mathbb{1}_{\{x_{t+h} \in R_m\}}}, \quad (3.12)$$

where  $T_v$  denotes the number of yearly observations in  $\Omega_v$ . Finding the optimal partitioning (split-variables and split-points) of the input space, in terms of minimizing squared error loss, poses a computationally infeasible combinatorial problem. The CART algorithm employs a greedy approach to finding the optimal partitioning of the input space, based on top-down induction [Breiman et al. 1984]. Note that if an explanatory variable does not contain any predictability, it is never optimal to split

on that variable. This is the sense in which the regression tree has built-in variable selection.

The depth of the regression tree controls the bias-variance tradeoff. To avoid overfitting, it is typical to impose a maximum depth restriction and require every leaf to contain a minimum number of observations (Hastie, Tibshirani, and Friedman [2009]; see the discussion below).

### 3.4.3.1 Gradient boosting

The gradient boosting algorithm of Friedman (2001) is based on the original ideas of boosting due to Kearns (1988); Valiant (1984); Freund and Schapire (1995), and seeks to reduce the bias of a single regression tree by growing multiple regression trees sequentially on the residuals from the preceding tree, then sum over their predictions.<sup>7</sup> The procedure is reflective of functional gradient descent; hence, the name of the algorithm.

Consider the general one-step ahead prediction model in (3.3) and the information set  $\Omega_v$ , characterized in Section 3.3. Gradient boosting assumes the function  $g$  in (3.3) is a sum of  $B$  regression trees:

$$y_{t+1|t+h} = \sum_{b=1}^B g_b(x_{t+h}), \quad (3.13)$$

where the regression trees  $g_b$  in (3.13) are grown sequentially. At the  $b$ -th step of the sequential tree-growing procedure, a new tree is grown such that it minimizes the objective function at that step, here taken to be the sum of squared errors:

$$J_b(\Theta_b) = \sum_{t=1}^{T_v} \left( y_{t+1} - \left[ g_{b-1}(x_{t+h}) + \sum_{m_b=1}^{M_b} c_{m_b} \mathbb{1}_{\{x_{t+h} \in R_{m_b}\}} \right] \right)^2, \quad (3.14)$$

where  $T_v$  denotes the number of yearly observations in  $\Omega_v$ ,  $\Theta_b$  characterizes the  $b$ -th regression tree in terms of split-variables, split points, and leaf values, and  $M_b$  denotes the number of disjoint subspaces in the partitioning of the  $b$ -th regression tree. When growing a new tree in order to minimize the objective function in (3.14), the solution is to grow the new tree based on the residuals from the preceding tree (Hastie et al. [2009]). The new tree is then added to the ensemble of existing trees according to a “gradient descent rule”:

$$g_b(x) = g_{b-1}(x) + \nu \sum_{m_b=1}^{M_b} c_{m_b} \mathbb{1}_{\{x \in R_{m_b}\}}, \quad (3.15)$$

where  $\nu \in (0, 1)$  determines the step size (or learning rate), and the tree predictions are analogous to the components of a negative gradient. Equation (3.15) is initialized by setting  $g_0(x)$  equal to the average of  $y$ .

---

<sup>7</sup>For objective functions more general than squared error loss, a generalized notion of residuals is considered (Hastie et al. [2009]).

To control the bias-variance tradeoff and to alleviate overfitting, I tune the number of trees in the ensemble  $B$ , the learning rate  $\nu$ , and the depth of the individual regression trees, using a validation sample; see Section 3.5.2. Since gradient boosting is inherently focused on bias reduction, the model is typically based on shallow regression trees with low variance (Hastie et al. 2009).

### 3.4.3.2 Random forest and bagging

The random forest algorithm of Breiman (2001) and the bootstrap aggregation (bagging) algorithm of Breiman (1996) seek to reduce the variance of a single regression tree by growing  $B$  regression trees on bootstrapped versions of the original data set, then average their predictions.<sup>8</sup> The Random forest algorithm is derived from bagging by a simple modification that aims to reduce correlation between the individual regression trees, as I discuss below.

Consider again the general one-step ahead prediction model in (3.3) and the information set  $\Omega_v$ , characterized in Section 3.3. Similar to gradient boosting, random forest and bagging assume that  $g$  in (3.3) can be represented as a sum of regression trees:

$$y_{t+1|t+h} = \frac{1}{B} \sum_{b=1}^B g_b(x_{t+h}). \quad (3.16)$$

The difference to gradient boosting is that the random forest and the bagging algorithm grow  $B$  regression trees independently on bootstrapped versions of the original data set.<sup>9</sup> The individual regression trees are, therefore, identically distributed, and the bias of the ensemble equals that of each individual tree. However, the variance of the ensemble is smaller than that of each individual tree, and the variance of the ensemble is decreasing in the number of trees in the ensemble  $B$ , and the correlation between the individual trees.<sup>10</sup> The random forest algorithm is derived from bagging by introducing a *dropout* mechanism in the tree-growing process to reduce correlation between the individual trees: only a subset  $\eta$  of the explanatory variables in  $x_{t+h}$ , selected randomly before each split point, are considered as candidates for split-variable in the random forest algorithm. By contrast, the bagging algorithm considers all explanatory variables in  $x_{t+h}$  as candidates for split variables at each split point.

---

<sup>8</sup>The bagging algorithm is sometimes formulated as a more general algorithm that can be used to fit other models than just regression trees to bootstrapped versions of the original data set (Breiman 1996).

<sup>9</sup>The random forest algorithm of Breiman (2001) and the bagging algorithm of Breiman (1996) rely on the paired bootstrap, which re-samples pairwise observations  $(y_{t+1}, x_{t+h})$  from the original data set with replacement. It is assumed that the leads and the lags of the explanatory variables in  $x_{t+h}$  capture relevant serial correlation in the data such that an explicit time series bootstrap, such as the block bootstrap, is not required. This seems to be a standard, albeit often tacit, assumption whenever the random forest or the bagging algorithm is applied to time series data (e.g., Gu et al. 2020).

<sup>10</sup>Recall that the variance of an average of  $B$  identically distributed random variables with variance  $\sigma^2$  and pairwise correlation  $\rho$  is  $\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2$ .

The random forest algorithm and the bagging algorithm are often celebrated for achieving hard-to-beat predictive performance, even when applied *off the shelf* [Hastie et al., 2009]. Therefore, I fix all hyperparameters of these algorithms at the values of the original Fortran implementation of random forest [Breiman and Cutler, 2004], see Table C.1 of the appendix, which is a common approach in the literature [Gu et al., 2020; Medeiros, Vasconcelos, Álvaro Veiga, and Zilberman, 2021].<sup>11</sup> In contrast to gradient boosting, there is no need to tune  $B$ , as larger values of  $B$  reduce variance without increasing bias. This is likely one of the main reasons that random forest and bagging are often considered immune to overfitting, albeit this claim is likely overly optimistic [Hastie et al., 2009]. Since the random forest and the bagging algorithm focus on variance reduction, they are typically based on relatively deep regression trees with low bias. I impose no maximum depth restriction, but require a minimum leaf size; see Table C.1 of the appendix.

### 3.4.4 Neural networks

Neural networks constitute a powerful class of nonparametric methods and have become almost synonymous with machine learning. Neural networks solve the curse of dimensionality primarily through feature extraction. Here, I focus on so-called feedforward neural networks.<sup>12</sup> By applying a linear regression model not directly to the explanatory variables, but to nonlinear combinations of the explanatory variables, derived through layers of functions composed together in a chain-like structure, feedforward neural networks allow for nonlinear effects of the explanatory variables on the variable of interest and interaction effects between the explanatory variables. Feedforward neural networks assume that  $g$  in (3.3) can be represented as

$$y_{t+1|t+h} = \kappa + \sum_{i=1}^{w_d} \gamma_i z_{it+h}^{(d)} \quad (3.17)$$

$$z_{j_d t+h}^{(d)} = \sigma \left( \kappa_{j_d}^{(d)} + \sum_{i_d=1}^{w_{d-1}} \Gamma_{i_d j_d}^{(d)} z_{i_d t+h}^{(d-1)} \right), \quad j_d = 1, \dots, w_d, \quad (3.18)$$

⋮

$$z_{j_2 t+h}^{(2)} = \sigma \left( \kappa_{j_2}^{(2)} + \sum_{i_2=1}^{w_1} \Gamma_{i_2 j_2}^{(2)} z_{i_2 t+h}^{(1)} \right), \quad j_2 = 1, \dots, w_2, \quad (3.19)$$

$$z_{j_1 t+h}^{(1)} = \sigma \left( \kappa_{j_1}^{(1)} + \sum_{i_1=1}^n \Gamma_{i_1 j_1}^{(1)} x_{i_1 t+h} \right), \quad j_1 = 1, \dots, w_1, \quad (3.20)$$

---

<sup>11</sup>Based on some unreported robustness checks, the results of the empirical application in Section 3.5 seem very insensitive to the choice of  $\eta$  and  $B$ .

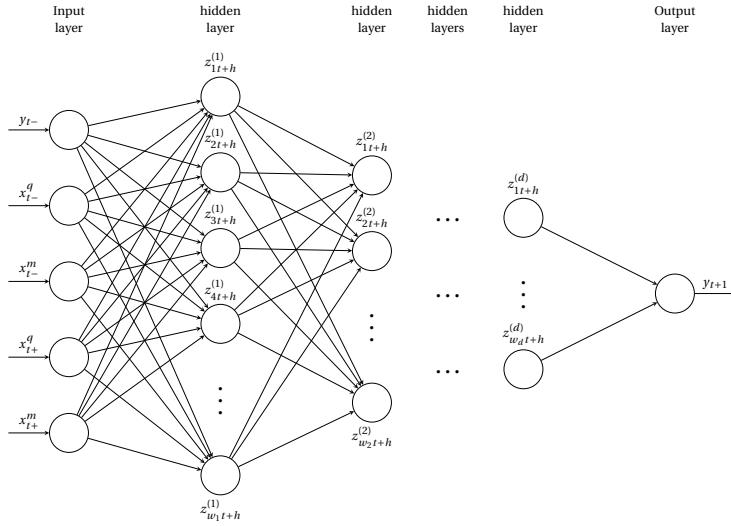
<sup>12</sup>Sometimes also referred to as multilayer perceptrons.

where  $n$  denotes the total number of distinct input signals in  $x_{t+h}$  (leads and lags of the explanatory variables). Note that (3.17) is a linear regression model in derived variables  $z_{t+h}^{(d)} \in \mathbb{R}^{w_d}$ , which constitute nonlinear transformations of the explanatory variables in  $x_{t+h}$ , learned through so-called *hidden layers* given by (3.20)-(3.18). In equation (3.17),  $\kappa$  is an intercept and  $\gamma \in \mathbb{R}^{w_d}$  is a column vector of slope coefficients. In the hidden layers (3.18)-(3.20),  $\kappa^{(l)} \in \mathbb{R}^{w_l}$  is a vector of intercepts and  $\Gamma^{(l)} \in \mathbb{R}^{w_l \times w_{l-1}}$  is a matrix of slope coefficients,  $l = 1, \dots, d$ . We follow the convention to set  $w_0 = n$ . The function  $\sigma$  is a so-called *activation function* and is used to induce nonlinearity into the model. I rely on the Swish activation function of Ramachandran et al. (2017), defined as  $\sigma(z) = z(1 + \exp(-z))^{-1}$  for  $z \in \mathbb{R}$ . The Swish activation function can be considered a smoothed version of the standard rectified linear unit (ReLU) activation function Glorot et al. (2011), defined as  $g(z) = \max(z, 0)$  for  $z \in \mathbb{R}$ , and is found to outperform the ReLU activation function on a number of different tasks (Ramachandran et al. 2017). I let  $d$  denote the number of hidden layers used to learn the derived variables, referred to as the *depth* of the model, and I let  $w_1, \dots, w_d$  denote the number of elements in each of the hidden layers, referred to as *widths*. The model is illustrated in Figure 3.2 by means of a directed acyclic graph (DAG).

The choice of depth  $d$  and widths  $w_1, \dots, w_d$  has implications for the bias-variance tradeoff. Deeper and wider networks tend to reduce bias at the expense of more variance (Goodfellow et al. 2016). However, it can be shown that even a shallow feed-forward network with just one hidden layer can approximate any Borel measurable function from one finite-dimensional space to another to any desired degree of accuracy provided the network is wide enough (Hornik et al. 1989, 1990, Cybenko 1989, Leshno et al. 1993). In practice, adding depth to the network often makes the model more difficult to optimize and more prone to overfitting (Goodfellow et al. 2016). To see the implications of different choices for the depth and the width parameters, I consider five different network “architectures” in the empirical application of the next section. Following Gu et al. (2020), I consider a network with a single hidden layer containing 32 units; a network with two hidden layers containing 32 units in the first layer and 16 units in the second layer; a network with three hidden layers containing 32 units in the first layer, 16 units in the second layer, and 8 units in the third layer; a network with four hidden layers containing 32 in the first layer, 16 units in the second layer, 8 units in the third layer, and 4 units in the fourth layer; and a network with five hidden layers containing 32 units in the first layer, 16 in the second layer, 8 units in the third layer, 4 in the fourth layer, and 2 units in the fifth layer.

It is evident from equations (3.17) - (3.20) that neural networks contain a lot of free parameters to estimate. Therefore, I employ five well-known regularization strategies to alleviate problems of overfitting: an  $L_1$  parameter norm penalization on the slope coefficients of the first hidden layer, dropout on the output of all hidden layers, learning rate shrinkage, early stopping, and ensembles.

I add an  $L^1$  parameter norm penalty on the slope coefficients of the first hidden

**Figure 3.2:** Feedforward neural network

Note: Illustration of the directed acyclic graph (DAG) associated with the feedforward neural network model in [3.17] - [3.20]. Edges represent how information travels through the model. The model is feedforward, as information only travels forward through the network without any feedback loops of information. Vertices of the input layer represent information that is presented to the model, vertices of the hidden layers represent elements of the sequentially derived  $z$ -variables, and the vertex of the output layer represent the model output.

layer to the sum of squared errors objective function. Note that, as the rest of the employed regularization strategies discourage sparsity, none of slope coefficients of the first hidden layer will be set exactly to zero in the empirical application of Section 3.5. However, based on some initial experimentation, I find that  $L_1$  regularization works to stabilize estimation. Assuming, as above, that we have  $T_v$  observations pairs available for estimation, the objective function used to estimate the parameters of the neural network model takes the form

$$J(\boldsymbol{\varkappa}, \boldsymbol{\Gamma}, \boldsymbol{\kappa}, \gamma) = \sum_{t=1}^{T_v} (y_{t+1} - \hat{y}_{t+1|t+h}(x_{t+h}))^2 + \xi \|\boldsymbol{\Gamma}^{(1)}\|, \quad (3.21)$$

where  $\hat{y}_{t+1|t+h}$  is represented by [3.17]-[3.20] and depends on the estimated parameters  $\hat{\boldsymbol{\varkappa}} \equiv (\hat{\boldsymbol{\kappa}}^{(1)}, \dots, \hat{\boldsymbol{\kappa}}^{(d)})$ ,  $\hat{\boldsymbol{\Gamma}} \equiv (\hat{\boldsymbol{\Gamma}}^{(1)}, \dots, \hat{\boldsymbol{\Gamma}}^{(d)})$ ,  $\hat{\boldsymbol{\kappa}}$ , and  $\hat{\boldsymbol{\gamma}}$ , and  $\xi$  is a hyperparameter that scales the contribution of the parameter norm penalty. I try different values of  $\xi$ ; see Table C.1 of the appendix.

I use gradient descent to minimize the objective function in (3.21) with respect to all free parameters simultaneously. It is often found to be important to shrink the step size of the gradient descent rule as the gradient approaches zero; otherwise, noise

in calculation of the gradient may dominate the directional signal (Gu et al. [2020]). I employ the adaptive moment estimation (Adam) variant of gradient descent due to Kingma and Ba (2014). Among other things, the Adam algorithm automatically shrinks individual learning rates for all free parameters during estimation. I keep all hyperparameters of the Adam algorithm at the suggested defaults Kingma and Ba (2014), and for each iteration of the Adam algorithm, I use all available observations to calculate the gradient, referred to as batch learning.

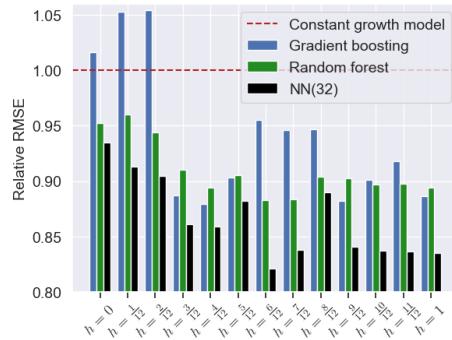
I apply the dropout regularization procedure of Hinton, Srivastava, Krizhevsky, Sutskever, and Salakhutdinov (2012) to the output of all hidden layers. For each iteration of the Adam algorithm, a new binary mask is randomly sampled and applied to the elements of the hidden layers. In this way, different elements of the hidden layers will be turned off for each iteration of the Adam algorithm. The mask for each element is sampled independently from all others, and the probability of sampling a mask value of one is a hyperparameter that is fixed a priori. I use the recommended value of 0.5 (Hinton et al. [2012]). The dropout procedure forces the neural network to learn derived variables that are “robust” and well-suited for out-of-sample analysis. Note that dropout can be interpreted as an approximation to bagging (Goodfellow et al. [2016]).

Early stopping seems to be the most popular form of regularization for neural networks and is motivated by the observation that errors measured over the estimation sample often decreases steadily during estimation, while errors measured over a separate validation sample often starts to increase again after a number of iterations. I use early stopping to stop the Adam algorithm once the loss measured over a separate validation sample has not decreased for a number of iterations, referred to as patience; see Section 3.5.2. Note that early stopping can be seen as an approximation to  $L_2$  regularization (Goodfellow et al. [2016]).

Finally, estimation of neural networks is often found to be sensitive to the choice of starting values used to initialize the numerical optimization routine (Goodfellow et al. [2016]). I estimate every neural network ten times, using different initializations, then construct an ensemble by averaging the predictions from the ten models with lowest validation loss; see Section 3.5.2. All intercepts are initialized to zero, and all slope coefficients are initialized by drawing from a truncated normal distribution as suggested by He et al. (2015).

### 3.5 Nowcasting experiment

This section presents a pseudo out-of-sample nowcasting experiment for U.S. CO<sub>2</sub> emissions. I make predictions  $y_{t+1|t+h} \equiv \mathbb{E}[y_{t+1} | \Omega_{t+h}]$ , where the information set  $\Omega_{t+h}$  is defined as in Section 3.4 for  $h \in \{0, \frac{1}{12}, \frac{2}{12}, \dots, 1\}$ . For  $h = 0$ , the predictions amount to one-year ahead forecasts using all available information from the previous year; for  $h = \frac{1}{12}$ , the predictions amount to nowcasts using all available information

**Figure 3.3:** Accuracy of the machine learning methods

Note: The figure displays out-of-sample, relative, root root mean squared errors for three of the most accurate machine learning methods, using the constant growth model as benchmark; “NN(32)” is a neural network model with one hidden layer containing 32 units.

from the previous year plus additional information available from the first month of the target year, and so forth. Note how this is in a sense a traditional forecasting experiment flipped on its head. Whereas a traditional forecasting experiment fixes the end-of-sample period and considers different forecast targets, I fix the forecast target and consider different end-of-sample periods.

Figure 3.3 summarizes the results of the experiment by reporting out-of-sample relative root mean squared errors for three of the most accurate machine learning methods across end-of-sample periods  $h$ , using the constant growth model as benchmark. The figure demonstrates that the machine learning methods can provide more accurate forecasts and nowcasts than the constant growth model, and that the machine learning methods seem capable of utilizing the stream of information that becomes available as we progress through the target year to produce repeatedly more accurate nowcasts.

### 3.5.1 Data

The data set used in this paper consists of an unbalanced panel of data sampled at mixed frequencies for the period 1960–2019. It contains yearly data on U.S. CO<sub>2</sub> emissions, monthly macroeconomic variables for the U.S., and quarterly macroeconomic variables for the U.S.

Data on U.S. CO<sub>2</sub> emissions consists of estimates from the Global Carbon Project [2020], which in turn is compiled from the inventories of the United Nations Framework Convention on Climate Change<sup>13</sup> (UNFCCC) and the Carbon Dioxide Informa-

<sup>13</sup> Accessible at [di.unfccc.int/time\\_series](https://di.unfccc.int/time_series), last accessed on August 24, 2021.

tion Analysis Center<sup>14</sup> (CIDAC). The CO<sub>2</sub> emissions estimates include emissions from fossil fuel combustion, oxidation, and cement production, and exclude emissions from bunker fuels, as the latter cannot be allocated unambiguously across countries. The series of U.S. CO<sub>2</sub> emissions is plotted in the top left panel of Figure 3.4. Note that CO<sub>2</sub> emissions are measured in meagatonnes (Mt), defined as 10<sup>6</sup> tonnes. At the time of writing, August 2021, official CO<sub>2</sub> emissions estimates are available only until 2019, which highlights the importance of being able to nowcast CO<sub>2</sub> emissions. Following Bennedsen et al. (2021c), total CO<sub>2</sub> emissions are transformed to per capita CO<sub>2</sub> emissions by dividing total emissions by the population size of the U.S. Data on the U.S. population is taken from the World Development Indicators database of the World Bank,<sup>15</sup> and is plotted in the top right panel of Figure 3.4. Per capita CO<sub>2</sub> emissions are plotted in the bottom left panel of Figure 3.4. Bennedsen et al. (2021c) find that U.S. per capita CO<sub>2</sub> emission contains a unit root, then suggest to take log-differences to achieve stationarity. I follow this suggestion. The series of log-differences of U.S. per capita emissions is plotted in the bottom right panel of Figure 3.4. Note this final series is an approximation to the growth rate of U.S. per capita CO<sub>2</sub> emissions, and that an initial observation is lost due to the stationarity transformation.

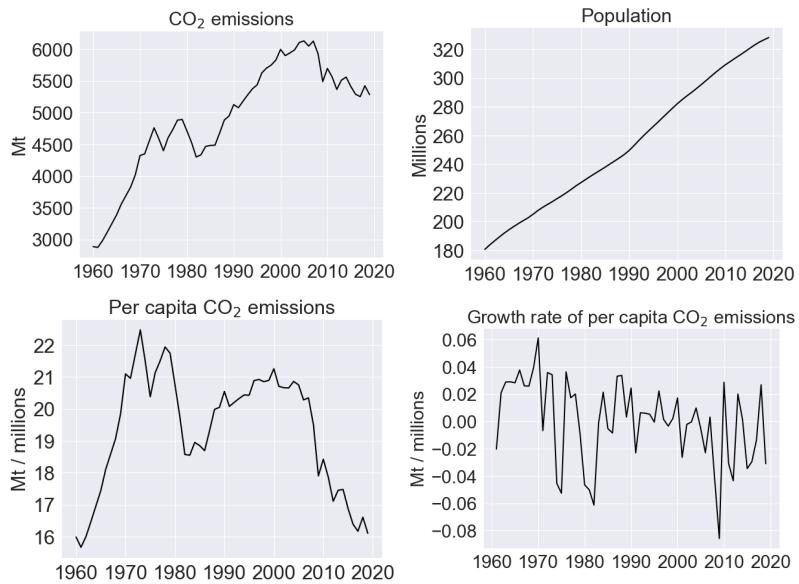
Data on monthly macroeconomic variables for the U.S. is from the “FRED-MD” database<sup>16</sup> put together by McCracken and Ng (2016). The database is well-suited for nowcasting, as it is updated and revised in real time: a new vintage of the database is released on the last business day of each month. In each new vintage, the most recent series are added, existing series are revised, and institutional changes to existing series are accounted for. Note that monthly series for, say, January, are not available until the February vintage, and some series are released with a longer delay, leading to a ragged edge. Unfortunately, real-time vintages of all series considered (CO<sub>2</sub> emissions, monthly macroeconomic variables, quarterly macroeconomic variables) are not available. Therefore, I construct pseudo-vintages of FRED-MD based on the vintage as of April, 2021, as I discuss below. The FRED-MD vintage as of April, 2021, contains 128 monthly variables; however, I remove 2 variables with no observations available for estimation in the initial pseudo vintage, leaving 126 monthly variables remaining.<sup>17</sup> Following McCracken and Ng (2016), these variables can be grouped into 8 distinct groups: output and income; labor market; housing; consumption, orders, and inventories; money and credit; interest and exchange rates; prices; and stock market. Further details on the set of monthly macroeconomic variables are provided in the appendix.

<sup>14</sup> Accessible at [cdiac.ess-dive.lbl.gov/data/](http://cdiac.ess-dive.lbl.gov/data/), last accessed on August 24, 2021.

<sup>15</sup> Accessible at [databank.worldbank.org/source/world-development-indicators](http://databank.worldbank.org/source/world-development-indicators). The population series “SPPOPTOTL” was downloaded on May 23, 2021.

<sup>16</sup> Accessible through Michael W. McCracken’s personal website: [research.stlouisfed.org/econ/mccracken/fred-databases/](http://research.stlouisfed.org/econ/mccracken/fred-databases/), last accessed on April 20, 2021.

<sup>17</sup> The monthly variables “ACOGNO” and “OILPRICEEx” have no observations available for estimation in the initial pseudo vintage; see the appendix for details.

**Figure 3.4:** United States carbon dioxide emissions

Note: In the bottom right panel, the growth rate of United States per capita CO<sub>2</sub> emissions is approximated by the log-difference of per capita CO<sub>2</sub> emissions; “Mt” is megatonne (i.e.  $10^6$  tonne).

Data on quarterly macroeconomic variables for the U.S. is from the “FRED-QD” database,<sup>18</sup> put together by McCracken and Ng (2020). Similar to the FRED-MD database, the FRED-QD database is updated and revised in real time, as discussed above. Note that quarterly series for, say, the first quarter, are not available until the April vintage, and some series are released with a longer delay, leading to a ragged edge. Similar to FRED-MD, I construct pseudo-vintages of FRED-QD based on the vintage as of April, 2021. The FRED-QD vintage as of April, 2021, contains 248 quarterly variables; however, I remove 111 variables from the FRED-QD database that are aggregations of variables in the FRED-MD database, which leaves 137 quarterly variables remaining. I remove 2 additional variables from FRED-QD with no observations available for estimation in the initial pseudo vintage, leaving 135 quarterly variables finally remaining.<sup>19</sup> Following McCracken and Ng (2020), these variables can be grouped into 14 distinct groups: national income and product accounts (NIPA); industrial production; employment and unemployment; housing; inventories, orders, and sales; prices; earnings and productivity; interest rates; money and credit; household and

<sup>18</sup> Accessible through Michael W. McCracken's personal website: [research.stlouisfed.org/econ/mccracken/fred-databases/](https://research.stlouisfed.org/econ/mccracken/fred-databases/), last accessed on April 20, 2021.

<sup>19</sup> The quarterly variables “SPCS20RSA” and “EXUSEU” have no observations available for estimation in the initial pseudo vintage; see the appendix for details.

balance sheets; exchange rates; other; stock markets; and non-household balance sheets. Further details on the set of quarterly macroeconomic variables are provided in the appendix.

I construct pseudo vintages of FRED-MD and FRED-QD, respectively, following an approach similar to Giannone et al. (2008), based on the simplifying assumption that the timing of data releases does not change over time. For each month from 1989 to 2019, I construct a pseudo vintage of FRED-MD by replicating the ragged edge of the FRED-MD vintage as of April, 2021<sup>20</sup>. Each pseudo vintage is also based on data from the FRED-MD vintage as of April, 2021. This approach, therefore, abstracts from keeping track of data revisions and institutional changes made to the data. For each quarter from 1989 to 2019, I construct a pseudo vintage of FRED-QD by replicating the ragged edge of the FRED-QD vintage as of April, 2021. Again, each pseudo vintage is also based on data from the FRED-QD vintage as of April, 2021. Note how this approach also imposes the implicit assumption that no additional, monthly series are released within a month, and that no additional, quarterly series are released within a quarter. For each pseudo vintage of FRED-MD and FRED-QD, respectively, all variables are transformed to stationarity and studentized to mean zero and unit variance following the procedure of McCracken and Ng (2016, 2020), and missing values are imputed following the approach discussed in Section 3.3.1.

### 3.5.2 Experimental setup

The nowcasting experiment is based on a rolling estimation window. Initially, I split the data set into three sub samples:

- An estimation sample, covering the years from 1960 to 1989.
- A validation sample, covering the years from 1990 to 1999.
- A test sample, covering the years from 2000 to 2019.

The estimation sample is used to estimate the models discussed in Section 3.4, the validation sample is used to tune the hyperparameters of the U-MIDAS model, the gradient boosting model, and the neural networks, and the test sample is reserved for out-of-sample comparison of predictive accuracy. I use exhaustive grid search to tune the hyperparameters of the U-MIDAS model, the gradient boosting model and the neural networks, based on a pre-specified grid of candidate values; see Table C.1 of the appendix. As I discuss below, the estimation sample and the validation sample are used differently for different models. Also, recall that model parameters are dependent on the forecast horizon (or equivalently, the end-of-sample period  $h$ ). Therefore, the same model needs to be estimated separately for  $h = 0$ ,  $h = 1/12$ ,

---

<sup>20</sup>The out-of-sample experiment is based on a rolling estimation window. I construct pseudo vintages from year 1989 onward, as 1989 corresponds to the last year of the initial estimation sample; see Section 3.5.2 for details.

$h = 2/12$ , and so forth, using a different pseudo vintage of the FRED-MD and the FRED-QD database.

For the U-MIDAS model and the gradient boosting model, the estimation sample, the validation sample, and the test sample are used as follows. For each end-of-sample period  $h$ , I estimate the U-MIDAS model and the gradient boosting model for every combination of candidate hyperparameter values specified in Table C.1 of the appendix, using only observations from the (initial) estimation sample (1960-1989), then use all models to predict the first observation in the validation sample (1990). That observation is then added to the end of the estimation sample, and all models are re-estimated using observations from the expanded estimation sample (1960-1990), then all models are used to predict the second observation in the validation sample (1991). I continue this procedure until the end of the validation sample. For the U-MIDAS model and the gradient boosting model, separately, the combination of hyperparameters that minimizes mean squared errors over the validation sample is considered optimal for making the first out-of-sample prediction. I collapse the estimation sample and the validation sample into a single sample of observations (1960-1999), and estimate the optimal U-MIDAS model and the optimal gradient boosting model on the collapsed sample of observations, then predict the first observation in the test sample (2000). Next, I *roll forward* the estimation sample and the validation sample so they cover the years 1961-1990 and 1991-2000, respectively, then use the above procedure to predict the next observation in the test sample (2001). I continue this procedure until the end of the test sample.

For the neural networks, I use the estimation sample, the validation sample, and test sample as follows. For each end of sample period  $h$ , I estimate the five neural network specifications described in Section 3.4.4 separately, for different candidate values of the  $L_1$  regularization hyperparameter  $\xi$ , specified in Table C.1 of the appendix, and ten different initializations, using observations from the estimation sample only (1960-1989). During estimation, early stopping is used to monitor the mean squared error loss over the validation sample (1990-1999), then stop the optimization routine once the validation loss no longer decreases. For each neural network specification, the ten models with lowest validation loss are used in an ensemble to predict the first observation in the test sample (2000), then average their predictions.<sup>21</sup> I roll forward the estimation sample and the validation sample as described above to predict the next observation in the test sample (2001). I continue this procedure until the end of the test sample.

For the univariate time series models, the MIDAS model, the random forest model, and the bagging model, I use the estimation sample, the validation sample, and test sample as follows. Initially, I collapse the estimation sample and the validation sample into a single sample of observations (1960-1999) used to estimate the models for each

---

<sup>21</sup>The ten models in the ensemble need not rely on the same choice of the  $L_1$  regularization hyperparameter  $\xi$ .

end-of-sample period  $h$ , then predict the first observation in the test sample (2000). Next, I roll forward the collapsed sample of observations used for estimation from covering the years 1960-1999 to covering the years 1961-2000, then predict the next observation in the test sample (2001). I continue this procedure until the end of the test sample.

For each end-of-sample period  $h$ , I use the root mean squared error (RMSE) loss metric to compare predictive accuracy of the different models over the test sample:

$$\text{RMSE}_h = \sqrt{\frac{1}{20} \sum_{t=2000}^{2019} (y_t - \hat{y}_{t|t-1+h})^2}.$$

I use the model confidence set proposed by Hansen, Lunde, and Nason [2011] to gauge statistically significant differences in RMSE across models. The model confidence set procedure determines a collection of models that contains the best model (here measured in terms of RMSE) with a given level of confidence. The model confidence set relies on a sequential testing procedure starting from the full set of models. An equivalence test is used to test if the “surviving” set of models are all equally accurate. If the equivalence test is rejected, inferior models are removed from the model confidence set using an elimination rule. I use the so-called  $T_R$  test statistic and  $e_R$  elimination rule; see Hansen et al. [2011] for details. I use the block bootstrap implementation of the model confidence set, and a block size of 12, as suggested by Hansen et al. [2011].

### 3.5.3 Experimental results

Table 3.1 reports the full set of results for the nowcasting experiment. Numbers reflect relative RMSEs, using the constant growth model as benchmark. Numbers less than one indicate that a given model is more accurate than the constant growth model. Asterisks indicate whether a particular model is included in the model confidence set: three asterisks indicate models in the 60% model confidence set; two asterisks indicate models in the 75% model confidence set; and one asterisk indicates models in the 90% model confidence set. The choice of confidence levels is inspired by Hansen et al. [2011]. Note that I construct separate model confidence sets for each end-of-sample period  $h$ . An appealing feature of the model confidence set is that it acknowledges the limitations of the data it is presented with. Uninformative data yields a model confidence set that contains a large number of models, informative data yields a model confidence set that contains only few models. This is evident from Table 3.1 where the number of models in the model confidence set is varying across  $h$ .

Since the univariate time series models do not include monthly or quarterly variables, they can be used only for forecasting. The RMSE of the these models is therefore constant across end-of-sample periods  $h$ . The univariate time series models

produce the least accurate forecasts of all models considered, and the random walk model produces the least accurate forecasts overall, which is perhaps not surprising given that the dependent variable is stationary.

In this experiment, I find the (restricted) MIDAS models to be very sensitive to the choice of starting values used for parameter optimization. I try multiple different starting values, but the performance of the MIDAS models does not appear robust across end of sample periods  $h^{22}$ . For most end-of-sample periods  $h$ , the MIDAS model with 10 macroeconomic input variable MIDAS(10) and the MIDAS model with 20 macroeconomic input variable MIDAS(20) perform worse than the constant growth model, and are rarely included in the model confidence set. However, for few values of  $h$ , MIDAS(10) and MIDAS(20) achieve a low RMSE. The MIDAS model with 30 macroeconomic input variables MIDAS(30) is never included in the model confidence set. Generally, the MIDAS model appears to perform gradually worse when including more explanatory variables. Comparing the performance of the MIDAS models to that of the U-MIDAS model, the latter appears more stable across  $h$ . Yet, the U-MIDAS model is never included in the model confidence set, and does not seem to outperform the constant growth model.

The U-MIDAS model, the tree-based methods, and the neural networks are presented with all input signals (leads and lags of the explanatory variables) available at the time of prediction. But, whereas the U-MIDAS model imposes a linear relationship between the input signals and emissions, the tree-based methods and the neural networks allow for nonlinearities and interaction effects. Note the total number of distinct input signals available at the time of prediction increases from 2,053 available for forecasting<sup>23</sup> ( $h = 0$ ) to 4,105 available for nowcasting by year's end<sup>24</sup> ( $h = 1$ ). The signal-to-noise ratio is, therefore, likely very low. In addition, the models have available only 40 observation pairs in the estimation sample and the validation sample, in total. Thus, the models appear to face a difficult learning task.

The tree-based methods achieve a reduction in RMSE between 4% and 12% relative to the constant growth model, across  $h$ . The methods seem capable of utilizing leads of the monthly and the quarterly macroeconomic variables that become available as we progress through the target year to produce nowcasts that are more accurate than the initial forecasts. Likely due to noise, the RMSE is not monotonically decreasing in  $h$ , but fluctuates around a clear decreasing trend. Interestingly, neither of the tree-based methods appear to benefit greatly from including leads of the macroeconomic variables beyond those available 3 to 4 months into the target year. For  $h < 3$ , the model confidence set suggests gradient boosting performs significantly worse than random forest and bagging at the confidence levels considered. This is

---

<sup>22</sup>I also experimented with constructing average predictions from multiple different initializations, but that did not robustify performance of the MIDAS models.

<sup>23</sup>One lagged dependent variable, 12 lags of the 126 monthly variables, and 4 lags of the 135 quarterly variables.

<sup>24</sup>An additional 12 leads of the 126 monthly variables and 4 leads of the 135 quarterly variables.

**Table 3.1:** Relative root mean squared errors for the pseudo out-of-sample experiment

	End of sample period $h$												
	0	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{4}{12}$	$\frac{5}{12}$	$\frac{6}{12}$	$\frac{7}{12}$	$\frac{8}{12}$	$\frac{9}{12}$	$\frac{10}{12}$	$\frac{11}{12}$	1
<i>Main benchmark model</i>													
Constant growth model	1.00	1.00**	1.00**	1.00	1.00	1.00	1.00	1.00	1.00	1.00*	1.00	1.00*	
<i>Univariate time series models</i>													
Random walk model	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	
AR(1)	1.07*	1.07*	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	
MA(1)	1.09*	1.09*	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	
ARMA(1, 1)	1.05*	1.05*	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	
<i>MIDAS regression models</i>													
MIDAS(10)	1.01***	1.24	<b>0.83***</b>	<b>0.77***</b>	0.90	1.03***	0.88	1.17	1.14	0.85***	1.06	1.17	1.08
<b>MIDAS(20)</b>	<b>0.83***</b>	1.19*	1.24	1.12	1.22	1.02***	1.11	1.51	1.19	1.43	1.48	1.12	1.26
MIDAS(30)	1.86	1.80	1.89	1.25	1.12	1.10	1.45	2.09	1.90	1.72	2.56	2.45	1.76
U-MIDAS	1.06	1.12	1.04	0.98	1.00	1.04	1.03	1.03	1.04	1.00	1.00	0.99	
<i>Tree-based models</i>													
Gradient boosting	1.02	1.05	1.05	0.89***	0.88	0.90***	0.96	0.95	0.95	0.88***	0.90*	0.92	0.89**
Bagging	0.96	0.98*	0.96	0.91***	0.92	0.89***	0.89	0.91	0.91***	0.89***	0.91	0.89**	0.90*
Random forest	0.95***	0.96**	0.94***	0.91***	0.89	0.90***	0.88	0.88	0.90***	0.90*	0.90	0.90**	0.89*
<i>Neural network models</i>													
NN(32)	0.93***	<b>0.91***</b>	0.90***	0.86***	0.86	0.88***	<b>0.82***</b>	<b>0.84***</b>	0.89***	<b>0.84***</b>	<b>0.84***</b>	<b>0.84***</b>	<b>0.83***</b>
NN(32, 16)	0.94***	0.91***	0.89***	0.87***	<b>0.83***</b>	<b>0.87***</b>	0.86	0.87	<b>0.87***</b>	0.86***	0.86*	0.85***	0.85***
NN(32, 16, 8)	0.93***	0.92***	0.90***	0.86***	0.90	0.88***	0.94	0.89	0.88***	1.04	0.92	0.92*	1.00
NN(32, 16, 8, 4)	0.97***	0.98*	1.12	0.96***	0.99	0.97***	0.98	1.02	0.97***	0.99	0.99	0.99	0.94*
NN(32, 16, 8, 4, 2)	1.01	1.00**	1.01	1.00	0.94	0.95***	1.06	0.95	1.01	1.00*	0.93*	0.98	0.97*

Note: Three asterisks indicate models in the 60% model confidence set; two asterisks indicate models in the 75% model confidence set; and one asterisk indicates models in the 90% model confidence set. Boldface numbers indicate models with lowest root mean squared error (RMSE) for a given end of sample period  $h$ ; "MIDAS(k)" refers to a MIDAS model with  $k$  input variables; and "NN(a,b,c,d,e)" refers to a neural network model with five hidden layers, containing  $a$  units in the first layer,  $b$  in the second layer,  $c$  in the third layer,  $d$  in the fourth layer, and  $e$  in the fifth layer.

despite extensive tuning of hyperparameters for gradient boosting and off-the-shelf implementations of random forest and bagging. For  $h < 3$ , random forest and bagging perform similarly, although random forest appears slightly preferred. For  $h \geq 3$ , the difference in performance between the tree-based methods is less profound. Overall, random forest seems to be the preferred method within the class of tree-based methods. It achieves the lowest RMSE in class for 7 of 13 values of  $h$  and is included in the model confidence set more often than the remaining tree-based methods at both the 75% and 90% confidence level. As discussed in Section 3.4.3, the central idea underlying gradient boosting, on the one hand, and random forest and bagging on the other hand, is different. Gradient boosting seeks to reduce the bias component of the mean squared error loss. Overfitting is an afterthought controlled by the choice of hyperparameters. Random forest and bagging are inherently focused on overfitting, and seek to reduce the variance component of the mean squared error loss. The difference between random forest and bagging is that random forest seeks to further reduce the variance component of the mean squared error loss through de-correlation of the individual regression trees that make up the ensemble. The preference for random forest emphasizes the importance of variance reduction, and that it is more important than bias reduction in this experiment. Given the large number of input signals, and the likely very low signal-to-noise ratio, it is perhaps not surprising that variance reduction seems more important than bias reduction.

The neural network models achieve a reduction in RMSE between 7% and 18% relative to the constant growth model, across  $h$ . Of all model classes considered, it seems neural networks perform the best, and that shallow networks with one or two hidden layers are preferred to deep networks with more hidden layers. For all confidence levels, the number of times the neural network model is included in the model confidence set, across  $h$ , is decreasing in the number of hidden layers. For networks with more than two hidden layers, performance across  $h$  appears unstable, similar to what we found for the MIDAS model, albeit not to the same degree. Shallow neural networks seem capable of utilizing leads of the monthly and the quarterly macroeconomic variables to produce nowcasts that are repeatedly more accurate than the initial forecasts. However, the benefit from including additional leads of the macroeconomic variables seems to decrease after some point. In particular, the networks do not appear to benefit much from including leads of the macroeconomic variables beyond those available halfway through the target year. The neural network model with a single hidden layer containing 32 units NN(32) and the neural network model with two hidden layers containing 32 units in the first hidden layer and 16 units in the second hidden layer NN(32,16), achieve the lowest RMSE overall for 10 of the 13 values of  $h$ , and either of the two models are included in the 60% model confidence set for all values of  $h$ . The model NN(32) seems to be the preferred model overall. It achieves the lowest RMSE overall for 7 of 13 values of  $h$ . It is included in the 60% model confidence set for 11 of 13 values of  $h$ , and included in the 75% model

confidence set for 12 of 13 values of  $h$ . Gu et al. (2020) find a similar preference for shallow networks when forecasting excess stock returns. As discussed in Section 3.4.4 deeper networks tend to reduce bias at the expense of more variance. The preference for shallow networks, therefore, reflects importance of variance reduction over bias reduction in much the same way did the preference for random forest over bagging and gradient boosting.

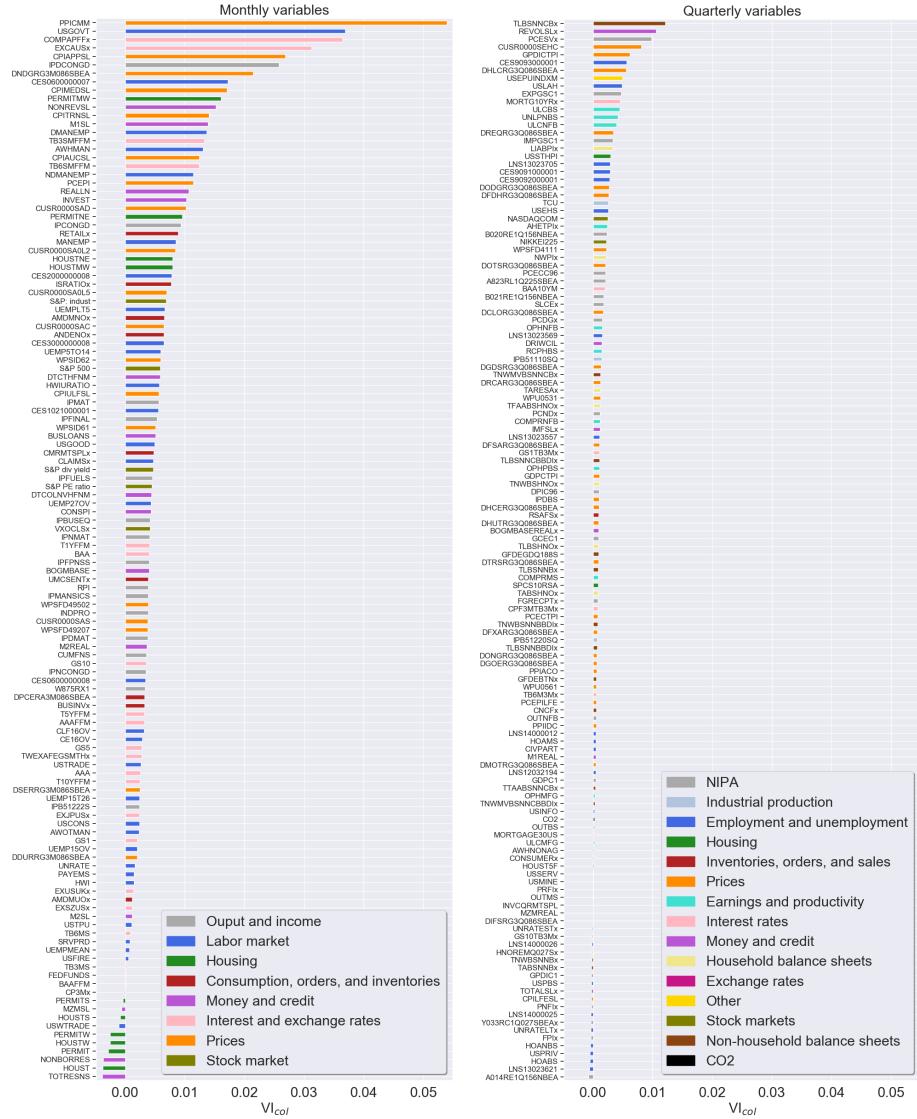
### 3.5.3.1 Variable importance

My aspirations toward interpreting the predictions of the machine learning methods are modest. I do not attempt to isolate the relative importance of nonlinearities, interactions effects, or the dimensionality reduction mechanisms embedded in the machine learning methods. Instead, I aim to quantify which of the input variables (defined below) that are the most important for the model output when controlling for the remaining input variables. I focus on the model NN(32), which was found to be the preferred model overall, and the end-of-sample period  $h = 1$ , for illustration. I construct a variable importance measure following Gu et al. (2020); Kelly, Pruitt, and Su (2019). The importance of a given input variable is defined by the reduction in  $R^2$  from setting all values of that input variable to zero while keeping remaining input variables at their observed levels and parameter estimates fixed.<sup>25</sup> I calculate two variants of this measure. First, all leads and lags of a given, explanatory variable are treated collectively as a single input variable. I denote this variant by  $VI_{col}$ . Next, each lead and lag of a given, explanatory variable is treated as a separate input variable. I denote this variant by  $VI_{sep}$ . Following Gu et al. (2020), I calculate  $VI_{col}$  and  $VI_{sep}$  consecutively within each estimation sample rolling forward through time, then average these into a single measure of  $VI_{col}$  and  $VI_{sep}$ . Finally, each measure is normalized to sum to one so that values reflect relative importance.

Figure 3.5 shows variable importance for the neural network specification NN(32) and end of sample period  $h = 1$ , based on the measure  $VI_{col}$ . The figure shows  $VI_{col}$  for the monthly macroeconomic variables in the plot on the left and  $VI_{col}$  for the quarterly macroeconomic variable and lagged CO<sub>2</sub> emissions in the plot on the right. Initially, we may note the measure is negative for few variables. According to  $VI_{col}$ , such variables contribute to overfitting, and the model would have benefited from omitting the variables. From Figure 3.5 it seems no variables are much more important than others. Therefore, the model likely relies on a combination of most variables in the data set for constructing its predictions. Figure 3.5 also color codes variables according to groups; see the appendix for details. The ten most important variables belong to either of the five groups prices, interest and exchange rates, labor market, output and income, and housing. However, the figure suggests that most

---

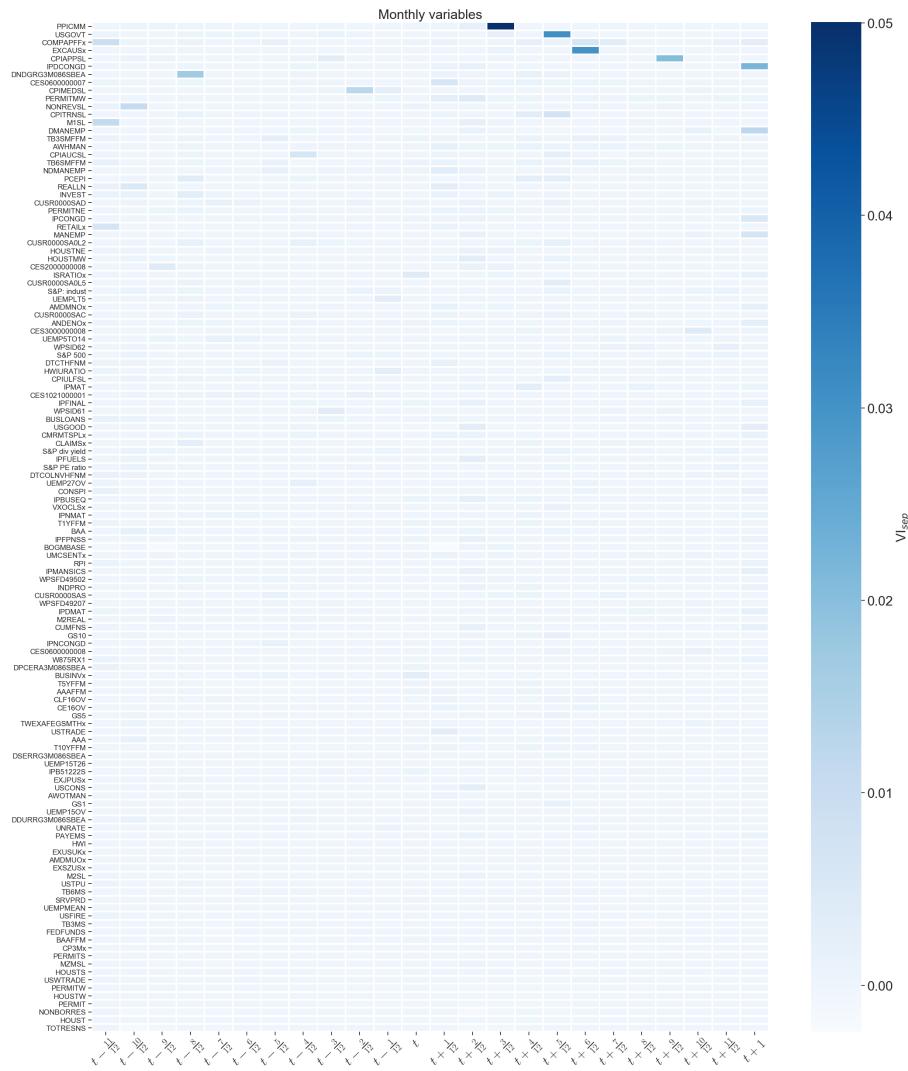
<sup>25</sup>Gu et al. (2020) find this simple measure behaves similarly to the variable importance measure of Dimopoulos, Bourret, and Lek (1995), which defines the importance of a given input variable by the sum of squared partial derivatives of the model with respect to that variable.

**Figure 3.5:** Variable importance for NN(32) and  $h = 1$ 

Note: Variables are named according to their FRED mnemonic and color coded according to groups; see the appendix for details. Lagged CO<sub>2</sub> emissions, named “CO2”, is included in the plot for quarterly variables. Negative values of VI<sub>col</sub> imply that variable contributes to overfitting.

groups are about equally important. Figure C.1 and C.2 of the appendix illustrates how VI<sub>col</sub> evolves over time for the monthly and the quarterly variables, respectively.

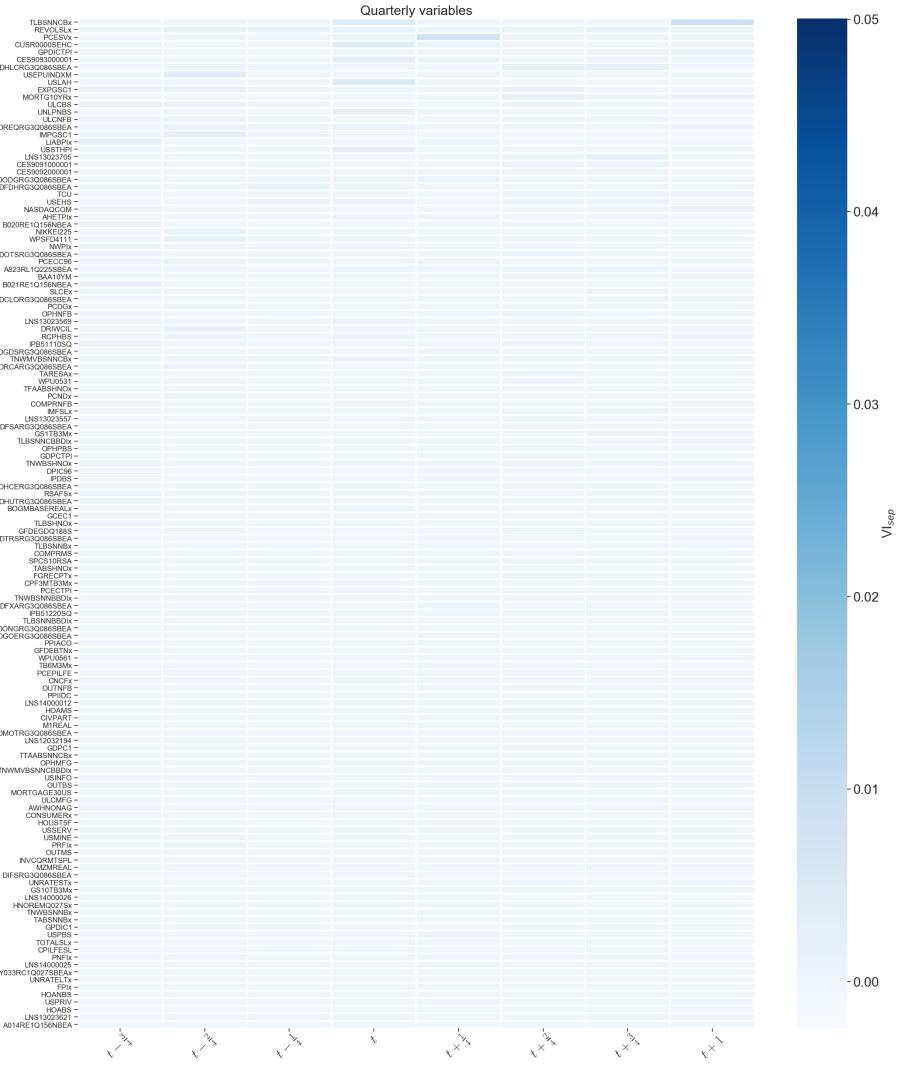
**Figure 3.6:** Importance of monthly macroeconomic variables across leads and lags for NN(32) and  $h = 1$



Note: Variables are named according to their FRED mnemonic; see the appendix for details. Variables are sorted as in Figure 3.5. Negative values of  $VI_{sep}$  imply that variable contributes to overfitting.

Values in Figures C.1 and C.2 have been scaled so they average to those in Figure 3.5. These figures suggest that  $\text{VI}_{col}$  is rather stable over time. As hinted in Section 3.4.4, the importance of variance reduction and the heavy regularization employed is

**Figure 3.7:** Importance of quarterly macroeconomic variables across leads and lags for NN(32) and  $h = 1$



Note: Variables are named according to their FRED mnemonic; see the appendix for details. Variables are sorted as in Figure 3.5. Lagged CO<sub>2</sub> emissions, named "CO2", is included in the plot. Negative values of  $VI_{sep}$  imply that variable contributes to overfitting.

probably what pushes the neural network model away from a sparse solution. Note that even though NN(32) is the simplest neural network specification considered, it contains a stunning number of free parameters to estimate. It contains a total of

80,161 free parameters to estimate for  $h = 0$ , and 131,425 free parameters to estimate for  $h = 1$ .<sup>26</sup> The model, therefore, requires heavy regularization to avoid overfitting.

From Figure 3.5 it seems the monthly macroeconomic variables are more important than the quarterly macroeconomic variables. However, note that if all distinct input signals were equally important, the monthly macroeconomic variables would appear the most important simply because they contain the most amount of distinct input signals (24 leads and lags in total for  $h = 1$ ). Figures 3.6 and 3.7 therefore, illustrate  $VI_{sep}$  across leads and lags for the monthly and the quarterly macroeconomic variables, respectively. Comparing the figures, gives an indication toward whether the importance of the monthly macroeconomic variables above the quarterly macroeconomic variables observed in Figure 3.5 seems to be a reflection of the former simply consisting of more distinct input signals; or, whether it appears that distinct monthly input signals are more important than the quarterly input signals. The figure seems to suggest that a combination of the two explanations may be true. For some of the monthly macroeconomic variables, we notice distinct leads and lags in Figure 3.6 that stand out as more important than any of the leads and lags of the quarterly macroeconomic variables in Figure 3.7. In particular, it seems the ordering of the monthly macroeconomic variables in the plot on the left in Figure 3.5 based on  $VI_{col}$ , is driven by few distinct leads and lags of the monthly macroeconomic variables. Interestingly, it is not always leads of the macroeconomic variables that are the most important. However, by comparing Figures 3.6 and 3.7 we also note that most distinct input signals appear about equally important for the monthly and the quarterly macroeconomic variables.

### 3.6 Conclusion

I investigate the use of machine learning methods for nowcasting the yearly growth rate of U.S. carbon dioxide ( $CO_2$ ) emissions over the period from 2000 to 2019, using a high-dimensional panel of macroeconomic variables sampled at mixed frequencies. As a special case, I also consider one-year ahead forecasting of U.S.  $CO_2$  emissions. To handle the problem of mixed frequencies, I propose to use the frequency alignment transformation from the mixed data sampling regression (MIDAS) literature to represent the data at a common frequency by transforming high-frequency variables into low-frequency vectors through skip-sampling (Ghysels et al. [2016]). The machine learning methods considered in this paper, tree-based methods (random forest, bagging, and gradient-boosting) and neural networks, rely on built-in variable selection mechanisms and feature extraction to solve the curse of dimensionality. They are also able to account for nonlinear effects of the macroeconomic variables on emissions and interaction effects between the macroeconomic variables.

---

<sup>26</sup>The input layer contains  $n \times 32$  slope coefficients and 32 intercepts, where  $n$  denotes the total number of distinct input signals. The output layer contains 32 slope coefficients and one intercept.

I find that machine learning methods are capable of utilizing the stream of macroeconomic information that becomes available as we progress through the target year to produce repeatedly more accurate nowcasts of U.S. CO<sub>2</sub> emissions that generally outperform forecasts from univariate time series models and nowcasts from MIDAS models. Neural networks deserve special attention, as they tend to produce the most accurate nowcasts, of the methods considered. In the experiment of this paper, shallow neural networks with only one or two hidden layers are preferred to deep neural networks with more hidden layers. Focusing on the particular case of a one-hidden-layer neural network and nowcasting by year's end, for illustration, I find that most input variables appear about equally important for the predictions of the model. However, the steps taken toward interpreting the predictions of the machine learning methods in this paper are modest. In future work, one could try to explain the sequence of revisions made to the initial forecast as we progress through the target year, and the relative importance of nonlinearities and interaction effects, by using model agnostic interpretation methods such as Shapley additive explanations (SHAP; Lundberg and Lee [2017]; Aas, Jullum, and Løland [2021]). Also, for the machine learning methods considered in this paper, the number of lags of the explanatory variables needs to be specified in advance. In future research, it could be interesting to consider the use of recurrent neural networks, like the long short-term memory (LSTM) network (Hochreiter and Schmidhuber [1997]; Gers et al. [1999]), which learns to decide the optimal number of lags of the explanatory variables by itself.

In essence, CO<sub>2</sub> emissions constitute a linear combination of different energy carriers, such as oil, coal, and natural gas, using deterministic, physical conversion factors (Marland and Rotty [1984]). Therefore, one could, alternatively, use the machine learning approach suggested in this paper to nowcast these energy carriers in a first step, then convert nowcasts of the energy carriers into nowcasts of CO<sub>2</sub> emissions in a second step. To the extent that energy consumption is more directly related to macroeconomic activity than CO<sub>2</sub> emissions themselves, this approach could potentially improve nowcasting of CO<sub>2</sub> emissions. I leave this approach for future research. The machine learning approach suggested in this paper could also be used to nowcast emissions of other types of greenhouse gases, emissions in other countries, or economic variables, such as GDP.

## Acknowledgements

I thank Professor Eric Hillebrand and Associate Professor Mikkel Bennedsen for useful comments and suggestions. I also acknowledge financial support from the DFF grant Econometric Modeling of Climate Change.

### 3.7 References

- Aas, K., Jullum, M., Løland, A., 2021. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence* 298, 103502.  
URL <https://www.sciencedirect.com/science/article/pii/S0004370221000539>
- Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., Veronese, G., 11 2010. New Eurocoin: Tracking Economic Growth in Real Time. *The Review of Economics and Statistics* 92 (4), 1024–1034.  
URL [https://doi.org/10.1162/REST\\_a\\_00045](https://doi.org/10.1162/REST_a_00045)
- Andreou, E., Ghysels, E., Kourtellos, A., 2010. Regression models with mixed sampling frequencies. *Journal of Econometrics* 158 (2), 246–261, specification Analysis in Honor of Phoebus J. Dhrymes.  
URL <https://www.sciencedirect.com/science/article/pii/S0304407610000072>
- Arrow et al., 1995. Economic growth, carrying capacity, and the environment. *Ecological Economics* 15 (2), 91 – 95.  
URL <http://www.sciencedirect.com/science/article/pii/0921800995000593>
- Auffhammer, M., Steinhauser, R., 2012. Forecasting The Path of U.S. CO<sub>2</sub> Emissions Using State-Level Information. *The Review of Economics and Statistics* 94 (1), 172–185.  
URL [https://doi.org/10.1162/REST\\_a\\_00152](https://doi.org/10.1162/REST_a_00152)
- Baffigi, A., Golinelli, R., Parigi, G., 2004. Bridge models to forecast the euro area GDP. *International Journal of Forecasting* 20 (3), 447–460.  
URL <https://www.sciencedirect.com/science/article/pii/S0169207003000670>
- Bai, J., Ghysels, E., Wright, J. H., 2013. State Space Models and MIDAS Regressions. *Econometric Reviews* 32 (7), 779–813.  
URL <https://doi.org/10.1080/07474938.2012.690675>
- Bai, J., Ng, S., 2002. Determining the Number of Factors in Approximate Factor Models. *Econometrica* 70 (1), 191–221.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00273>
- Bai, J., Ng, S., 2008. Forecasting economic time series using targeted predictors. *Journal of Econometrics* 146 (2), 304–317, honoring the research contributions of Charles R. Nelson.  
URL <https://www.sciencedirect.com/science/article/pii/S0304407608001085>
- Banbura, M., Giannone, D., Reichlin, L., 2010. Nowcasting.

- Ba  nbara, M., Modugno, M., 2014. Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics* 29 (1), 133–160.  
 URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2306>
- Bennedsen, M., Hillebrand, E., Jensen, S., 2021a. A Neural Network Approach to the Environmental Kuznets Curve, Chapter 1 of this dissertation.
- Bennedsen, M., Hillebrand, E., Jensen, S., 2021b. Apocalypse Now? Projecting CO<sub>2</sub> Emissions with Neural Networks, Chapter 2 of this dissertation.
- Bennedsen, M., Hillebrand, E., Koopman, S. J., 2021c. Modeling, forecasting, and nowcasting U.S. CO<sub>2</sub> emissions using many macroeconomic predictors. *Energy Economics* 96, 105118.  
 URL <https://www.sciencedirect.com/science/article/pii/S0140988321000232>
- Bouwman, A., Kram, T., Klein Goldewijk, K., 2006. Integrated modelling of global environmental change: an overview of IMAGE 2.4. Netherlands Environmental Assessment Agency (MNP), Bilthoven, The Netherlands.
- Breiman, L., 1996. Bagging predictors. *Machine learning* 24 (2), 123–140.
- Breiman, L., 2001. Random forests. *Machine learning* 45 (1), 5–32.
- Breiman, L., Cutler, A., 2004. Random Forests. Fortran code version 5.1, available for download at [https://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_software.htm](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_software.htm).
- Breiman, L., Friedman, J., Stone, C., Olshen, R., 1984. Classification and Regression Trees. Taylor & Francis.  
 URL <https://books.google.dk/books?id=JwQx-W0mSyQC>
- Calvin, K., 2011. The global Change Assement Model. <http://wiki.umd.edu/gcam/>.
- Chamberlain, G., Rothschild, M., 1983. Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica* 51 (5), 1281–1304.  
 URL <http://www.jstor.org/stable/1912275>
- Dimopoulos, Y., Bourret, P., Lek, S., 1995. Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters* 2 (6), 1–4.  
 URL <http://dblp.uni-trier.de/db/journals/npl/npl2.html#DimopoulosBL95>
- Dinda, S., 2004. Environmental Kuznets Curve Hypothesis: A Survey. *Ecological Economics* 49 (4), 431 – 455.  
 URL <https://www.sciencedirect.com/science/article/pii/S0921800904001570>

- Diron, M., 2008. Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data. *Journal of Forecasting* 27 (5), 371–390.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/for.1067>
- Doz, C., Giannone, D., Reichlin, L., 2011. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics* 164 (1), 188–205, annals Issue on Forecasting.  
URL <https://www.sciencedirect.com/science/article/pii/S030440761100039X>
- Emmerling et al., 2016. The WITCH 2016 Model - Documentation and Implementation of the Shared Socioeconomic Pathways. Tech. rep., Fondazione Eni Enrico Mattei (FEEM).  
URL <http://www.jstor.org/stable/resrep15162.4>
- Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The Generalized Dynamic-Factor Model: Identification and Estimation. *The Review of Economics and Statistics* 82 (4), 540–554.  
URL <http://www.jstor.org/stable/2646650>
- Foroni, C., Marcellino, M., Schumacher, C., 2015. Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 178 (1), 57–82.  
URL <http://www.jstor.org/stable/43965717>
- Freund, Y., Schapire, R. E., 1995. A desicion-theoretic generalization of on-line learning and an application to boosting. In: Vitányi, P. (Ed.), *Computational Learning Theory*. Springer Berlin Heidelberg, Berlin, Heidelberg, 23–37.
- Friedman, J. H., 2001. Greedy function approximation: A gradient boostingmachine. *The Annals of Statistics* 29 (5), 1189 – 1232.  
URL <https://doi.org/10.1214/aos/1013203451>
- Fujimori, S., Masui, T., Yuzuru, M., 2012. AIM/CGE [basic] manual. Discussion paper series. Tsukuba, Japan: Center for Social and Environmental Systems Research, NIES.
- Gers, F., Schmidhuber, J., Cummins, F., 1999. Learning to forget: continual prediction with LSTM. In: 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470). Vol. 2. 850–855 vol.2.
- Ghysels, E., 2016. MIDAS Matlab Toolbox.
- Ghysels, E., Kvedaras, V., Zemlys, V., 2016. Mixed frequency data sampling regression models: the R package midasr. *Journal of statistical software* 72 (1), 1–35.

- Ghysels, E., Santa-Clara, P., Valkanov, R., 2004. The MIDAS Touch: Mixed Data Sampling Regression Models. Cirano working papers, CIRANO.  
 URL <https://EconPapers.repec.org/RePEc:cir:cirwor:2004s-20>
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return trade-off after all. Journal of Financial Economics 76 (3), 509–548.  
 URL <https://www.sciencedirect.com/science/article/pii/S0304405X05000140>
- Ghysels, E., Sinko, A., Valkanov, R., 2007. MIDAS Regressions: Further Results and New Directions. Econometric Reviews 26 (1), 53–90.  
 URL <https://doi.org/10.1080/07474930600972467>
- Giannone, D., Reichlin, L., Small, D., 2008. Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics 55 (4), 665–676.  
 URL <https://www.sciencedirect.com/science/article/pii/S0304393208000652>
- Global Carbon Project, 2020. Supplemental data of Global Carbon Budget 2020 (Version 1.0) [Data set]. Global Carbon Project.
- Glorot, X., Bordes, A., Bengio, Y., 11–13 Apr 2011. Deep Sparse Rectifier Neural Networks. In: Gordon, G., Dunson, D., Dudík, M. (Eds.), Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Vol. 15 of Proceedings of Machine Learning Research. PMLR, Fort Lauderdale, FL, USA, 315–323.  
 URL <http://proceedings.mlr.press/v15/glorot11a.html>
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, <http://www.deeplearningbook.org>.
- Grossman, G. M., Krueger, A. B., November 1991. Environmental Impacts of a North American Free Trade Agreement. Working Paper 3914, National Bureau of Economic Research.  
 URL <http://www.nber.org/papers/w3914>
- Gu, S., Kelly, B., Xiu, D., 02 2020. Empirical Asset Pricing via Machine Learning. The Review of Financial Studies 33 (5), 2223–2273.  
 URL <https://doi.org/10.1093/rfs/hhaa009>
- Hansen, P. R., Lunde, A., Nason, J. M., 2011. THE MODEL CONFIDENCE SET. Econometrica 79 (2), 453–497.  
 URL <http://www.jstor.org/stable/41057463>
- Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference and prediction, 2nd Edition. Springer.  
 URL <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>

- He, K., Zhang, X., Ren, S., Sun, J., December 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: The IEEE International Conference on Computer Vision (ICCV).
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. R., 2012. Improving neural networks by preventing co-adaptation of feature detectors.
- Hochreiter, S., Schmidhuber, J., 11 1997. Long Short-Term Memory. Neural Computation 9 (8), 1735–1780.  
URL <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hoerl, A. E., Kennard, R. W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12 (1), 55–67.  
URL <https://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634>
- Holtz-Eakin, D., Selden, T. M., 1995. Stoking the fires? CO<sub>2</sub> emissions and economic growth. Journal of Public Economics 57 (1), 85 – 101.  
URL <http://www.sciencedirect.com/science/article/pii/004727279401449X>
- IPCC, 2014. AR5 Synthesis Report: Climate Change 2014.
- Kearns, M., Dec. 1988. Thoughts on Hypothesis Boosting, unpublished.
- Kelly, B. T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134 (3), 501–524.  
URL <https://www.sciencedirect.com/science/article/pii/S0304405X19301151>
- Kingma, D. P., Ba, J., 2014. Adam: A Method for Stochastic Optimization.
- Luderer et al., 2013. Description of the REMIND Model (Version 1.5). Available at SSRN: <https://ssrn.com/abstract=2312844> or <http://dx.doi.org/10.2139/ssrn.2312844>.
- Lundberg, S. M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st international conference on neural information processing systems. 4768–4777.
- Marcellino, M., Schumacher, C., 2010. Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP\*. Oxford Bulletin of Economics and Statistics 72 (4), 518–550.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0084.2010.00591.x>
- Marland, G., Rotty, R. M., 1984. Carbon dioxide emissions from fossil fuels: a procedure for estimation and results for 1950–1982. Tellus B: Chemical and Physical Meteorology 36 (4), 232–261.  
URL <https://doi.org/10.3402/tellusb.v36i4.14907>

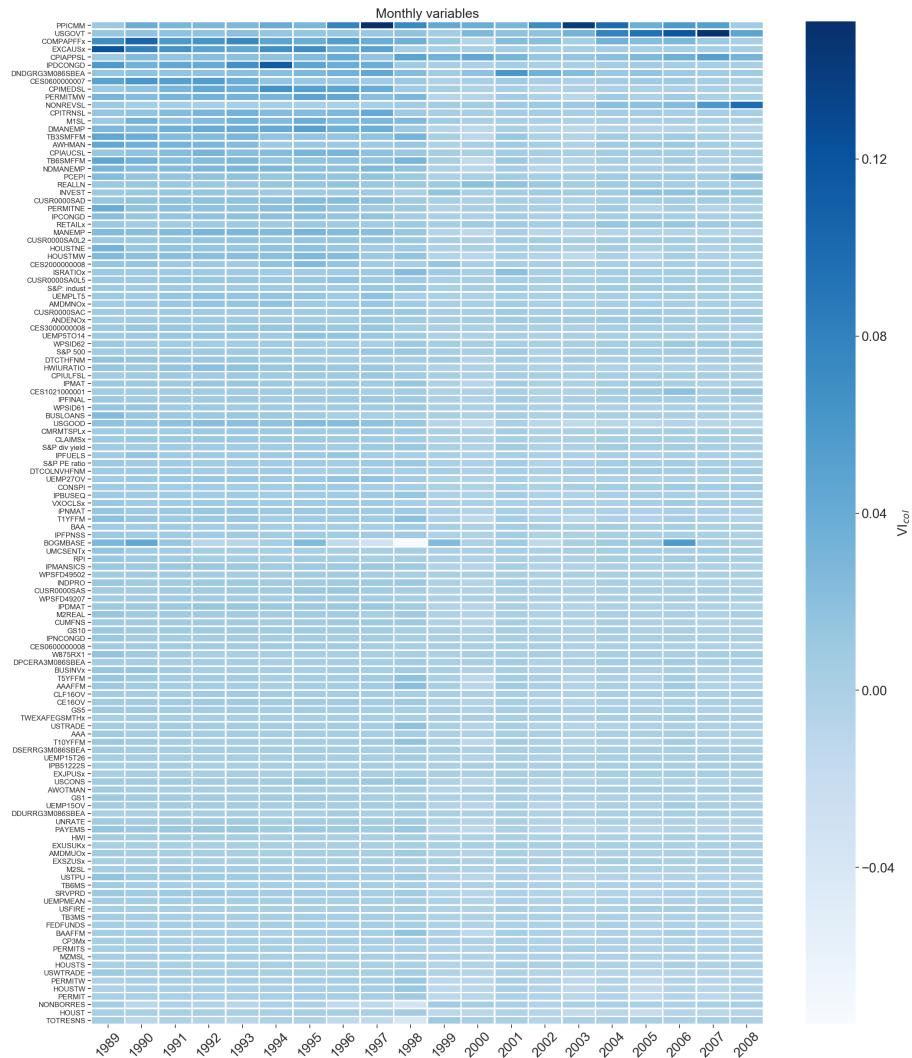
- McCracken, M., Ng, S., March 2020. FRED-QD: A Quarterly Database for Macroeconomic Research. Working Paper 26872, National Bureau of Economic Research.  
URL <http://www.nber.org/papers/w26872>
- McCracken, M. W., Ng, S., 2016. FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics* 34 (4), 574–589.  
URL <https://doi.org/10.1080/07350015.2015.1086655>
- Medeiros, M. C., Vasconcelos, G. F. R., Álvaro Veiga, Zilberman, E., 2021. Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business & Economic Statistics* 39 (1), 98–119.  
URL <https://doi.org/10.1080/07350015.2019.1637745>
- Messner, S., Strubegger, M., July 1995. User's Guide for MESSAGE III. IIASA Working Paper, IIASA, Laxenburg, Austria.  
URL <http://pure.iiasa.ac.at/id/eprint/4527/>
- Ramachandran, P., Zoph, B., Le, Q. V., 2017. Searching for Activation Functions.
- Schmalensee, R., Stoker, T., Judson, R., 02 1998. World Carbon Dioxide Emissions: 1950-2050. *The Review of Economics and Statistics* 80, 15–27.
- Simon, N., Friedman, J., Hastie, T., Tibshirani, R., 2013. A Sparse-Group Lasso. *Journal of Computational and Graphical Statistics* 22 (2), 231–245.  
URL <http://www.jstor.org/stable/43304828>
- Stern, D. I., Apr 2017. The environmental Kuznets curve after 25 years. *Journal of Bioeconomics* 19 (1), 7–28.  
URL <https://doi.org/10.1007/s10818-017-9243-1>
- Stock, J. H., Watson, M. W., 2002. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* 97 (460), 1167–1179.  
URL <http://www.jstor.org/stable/3085839>
- Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1), 267–288.  
URL <http://www.jstor.org/stable/2346178>
- United Nations, 2015. Paris Agreement. Adopted on 12 December 2015 and entered into force on 4 November 2016.
- Valiant, L. G., Nov. 1984. A Theory of the Learnable. *Commun. ACM* 27 (11), 1134–1142.  
URL <https://doi.org/10.1145/1968.1972>

- Wagner, M., 2008. The carbon Kuznets curve: A cloudy picture emitted by bad econometrics? *Resource and Energy Economics* 30 (3), 388 – 408.  
URL <http://www.sciencedirect.com/science/article/pii/S0928765507000504>
- Wagner, M., 2015. The Environmental Kuznets Curve, Cointegration and Nonlinearity. *Journal of Applied Econometrics* 30 (6), 948–967.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2421>
- Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B* 68, 49–67.
- Zhao, X., Du, D., 2015. Forecasting carbon dioxide emissions. *Journal of Environmental Management* 160, 39 – 44.  
URL <http://www.sciencedirect.com/science/article/pii/S0301479715300931>
- Zou, H., Hastie, T., 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67 (2), 301–320.  
URL <http://www.jstor.org/stable/3647580>

## Appendix

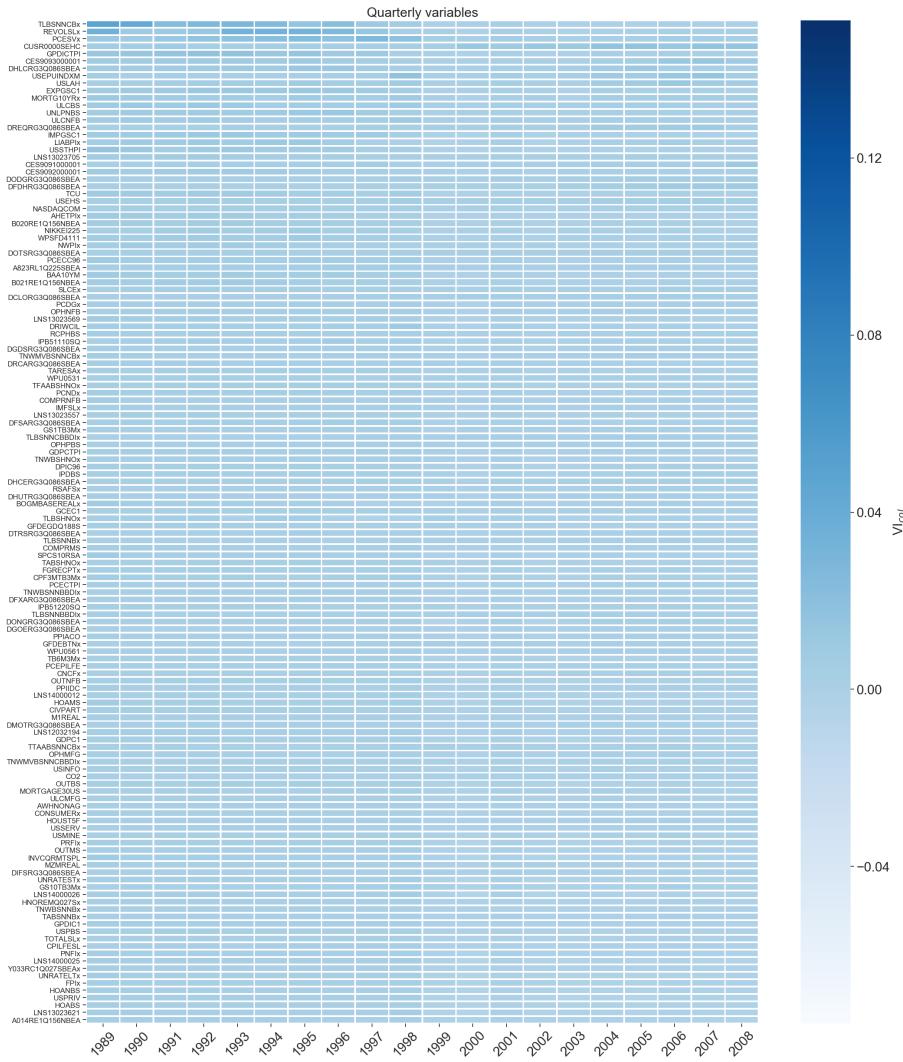
### C.1 Figures

**Figure C.1:** Importance of monthly macroeconomic variables over time for NN(32) and  $h = 1$



Note: Variables are named according to their FRED mnemonic; see Appendix C.3 and C.4 for details. Variables are sorted as in Figure 3.5. Years on the bottom axis refer to the last year of the estimation sample used to calculate  $VI_{col}$ . Values of  $VI_{col}$  are scaled so they average to those in Figure 3.5. Negative values of  $VI_{col}$  imply that variable contributes to overfitting.

**Figure C.2:** Importance of quarterly macroeconomic variables over time for NN(32) and  $h = 1$



Note: Variables are named according to their FRED mnemonic; see Appendix C.3 and C.4 for details. Variables are sorted as in Figure 3.5. Lagged CO<sub>2</sub> emissions, named "CO2", is included in the plot. Years on the bottom axis refer to the last year of the estimation sample used to calculate VI<sub>col</sub>. Values of VI<sub>col</sub> are scaled so they average to those in Figure 3.5. Negative values of VI<sub>col</sub> imply that variable contributes to overfitting.

## C.2 Tables

**Table C.1:** Machine learning hyperparameters

Gradient boosting	Random Forest	Bagging	Neural network models
Min. leaf size = 5	Min. leaf size = 5	Min. leaf size = 5	$\xi^* \in \{10^{-5}, 10^{-4}, 10^{-3}\}$
Max. tree depth* $\in \{1, 2, 3\}$	$B = 500$	$B = 500$	Dropout-rate = 0.5
$B^* \in \{50, 100, \dots, 1,000\}$	$\eta = \lfloor n/3 \rfloor$	$\eta = 1$	Ensemble size = 10
$\nu^* \in \{10^{-3}, 10^{-2}, 10^{-1}\}$			Patience = 1,000 Activation function: Swish Initializer: He normal Optimizer: Adam

Note: Asterisks indicate hyperparameters that are tuned using the validation sample, described in the main text; “n” is used to denote the total number of distinct input signals (leads and lags of the explanatory variables); all hyperparameters of the Adam algorithm are at the suggested defaults of Kingma and Ba (2014).

For the U-MIDAS model, I sample 1,000 candidate values of the  $L^1$  regularization parameter  $\lambda$  evenly on a logarithmic scale (a geometric progression) from  $[10^{-4}, 10^{-1}]$ .

**Table C.2:** Adam algorithm

---

**Require:** Step size  $\epsilon$  (0.001)  
**Require:** Exponential decay rates  $\rho_1, \rho_2 \in [0, 1]$  for moment estimates (0.9, 0.999)  
**Require:** Small constant  $\delta$  used for numerical stabilization ( $10^{-8}$ )  
**Require:** Initial parameters  $\theta$

Initialize first moment vector:  $m = 0$   
 Initialize second moment vector:  $v = 0$   
 Initialize time step:  $t = 0$

**while** stopping criterion not met **do**

- $t \leftarrow t + 1$
- Compute gradient:  $g \leftarrow \nabla_{\theta} J(\theta)$
- Update biased first moment estimate:  $m \leftarrow \rho_1 m + (1 - \rho_1)g$
- Update biased second moment estimate:  $v \leftarrow \rho_2 v + (1 - \rho_2)g \odot g$
- Correct bias in first moment:  $\hat{m} \leftarrow m / (1 - \rho_1^t)$
- Correct bias in second moment:  $\hat{v} \leftarrow v / (1 - \rho_2^t)$
- compute update:  $\Delta\theta = -\epsilon \cdot \hat{m} / (\sqrt{\hat{v}} + \delta)$
- Apply update:  $\theta \leftarrow \theta + \Delta\theta$

**end while**

---

Note: Numbers in parenthesis are suggested defaults.

The Adam algorithm individually adapts the learning rate of all parameters in two ways. First, its scales them inversely proportional to the square root of an exponentially decaying average of past squared values of the gradient (second moment estimate). In this way, the learning rate of parameters with small partial derivatives of the cost function decreases less rapidly than that of parameters with large partial derivatives. This implies greater progress in more gently sloped regions of the parameter space. Second, to speed up optimization, especially in face of pathological curvature, Adam incorporates momentum by scaling the individual learning rates proportionally to an exponentially decaying average of past values of the gradient (first moment estimate). Finally, Adam includes bias corrections of the moment estimates to account for initialization at the origin.

### C.3 Quarterly data description

Data on quarterly macroeconomic variables is from the "FRED-QD" database, put together by McCracken and Ng (2020). This paper relies on data from the FRED-QD vintage as of April, 2021. Following McCracken and Ng (2020), the data in FRED-QD can be grouped into 14 distinct groups, provided in the tables below, which are adapted from McCracken and Ng (2020): "column ID" denotes series number; "Tcode" denotes one of the following stationarity transformations: (1) no transformation; (2)  $\Delta x_{t_q}$ ; (3)  $\Delta^2 x_{t_q}$ ; (4)  $\log x_{t_q}$ ; (5)  $\Delta \log x_{t_q}$ ; (6)  $\Delta^2 \log x_{t_q}$ ; (7)  $\Delta(x_{t_q} / x_{t_q-1} - 1.0)$ ; "MD" denotes whether a series is an aggregation of a series from FRED-MD (1: true; 0: false); "FRED mnemonic" is mnemonics used by the Federal Reserve Bank of St. Louis; and "FRED are descriptions" are descriptions used by the Federal Reserve Bank of St. Louis. An asterisk on ID-number is used to indicate series with no observations available for estimation in the initial pseudo vintage, as discussed in the main text.

**Table C.3:** Group 1: NIPA

ID	Tcode	MD	FRED mnemonic	FRED Description
1	1	5	0	GDPC1
2	2	5	0	PCECC96
3	3	5	0	PCDGx
4	4	5	0	PCESVx
5	5	5	0	PCNDx
6	6	5	0	GPDIC1
7	7	5	0	FPIx
8	8	5	0	Y033RC1Q027SBEAx

9	9	5	0	PNFIx	Real private fixed investment: Nonresidential (Billions of Chained 2012 Dollars), deflated using PCE
10	10	5	0	PRFIx	Real private fixed investment: Residential (Billions of Chained 2012 Dollars), deflated using PCE
11	11	1	0	A014RE1Q156NBEA	Shares of gross domestic product: Gross private domestic investment: Change in private inventories (Percent)
12	12	5	0	GCEC1	Real Government Consumption Expenditures & Gross Investment (Billions of Chained 2012 Dollars)
13	13	1	0	A823RL1Q225SBEA	Real Government Consumption Expenditures and Gross Investment: Federal (Percent Change from Preceding Period)
14	14	5	0	FGRECPTx	Real Federal Government Current Receipts (Billions of Chained 2012 Dollars), deflated using PCE
15	15	5	0	SLCEx	Real government state and local consumption expenditures (Billions of chained 2012 Dollars), deflated using PCE
16	16	5	0	EXPGSC1	Real Exports of Goods & Services, 3 Decimal (Billions of Chained 2012 Dollars)
17	17	5	0	IMPGSC1	Real Imports of Goods & Services, 3 Decimal (Billions of Chained 2012 Dollars)
18	18	5	0	DPIC96	Real Disposable Personal Income (Billions of Chained 2012 Dollars)
19	19	5	0	OUTNFB	Nonfarm Business Sector: Real Output (Index 2012=100)
20	20	5	0	OUTBS	Business Sector: Real Output (Index 2012=100)
21	21	5	0	OUTMS	Manufacturing Sector: Real Output (Index 2012=100)
22	190	2	0	B020RE1Q156NBEA	Shares of gross domestic product: Exports of goods and services (Percent)
23	191	2	0	B021RE1Q156NBEA	Shares of gross domestic product: Imports of goods and services (Percent)

**Table C.4:** Group 2: Industrial production

	ID	Tcode	MD	FRED mnemonic	FRED Description
1	22	5	1	INDPRO	Industrial Production Index (Index 2012=100)
2	23	5	1	IPFINAL	Industrial Production: Final Products (Market Group) (Index 2012=100)
3	24	5	1	IPCONGD	Industrial Production: Consumer Goods (Index 2012=100)
4	25	5	1	IPMAT	Industrial Production: Materials (Index 2012=100)
5	26	5	1	IPDMAT	Industrial Production: Durable Materials (Index 2012=100)
6	27	5	1	IPNMAT	Industrial Production: Nondurable Materials (Index 2012=100)
7	28	5	1	IPDCONGD	Industrial Production: Durable Consumer Goods (Index 2012=100)
8	29	5	0	IPB51110SQ	Industrial Production: Durable Goods: Automotive products (Index 2012=100)
9	30	5	1	IPNCONGD	Industrial Production: Nondurable Consumer Goods (Index 2012=100)
10	31	5	1	IPBUSEQ	Production: Business Equipment (Index 2012=100)
11	32	5	0	IPB51220SQ	Industrial Production: Consumer energy products (Index 2012=100)
12	33	1	0	TCU	Capacity Utilization: Total Industry (Percent of Capacity)
13	34	1	1	CUMFNS	Capacity Utilization: Manufacturing (SIC) (Percent of Capacity)
14	194	5	1	IPMANSICS	Industrial Production: Manufacturing (SIC) (Index 2012=100)
15	195	5	1	IPB51222S	Industrial Production: Residential Utilities (Index 2012=100)
16	196	5	1	IPFUELS	Industrial Production: Fuels (Index 2012=100)

**Table C.5:** Group 3: Employment and unemployment

	ID	Tcode	MD	FRED mnemonic	FRED Description
1	35	5	1	PAYEMS	All Employees: Total nonfarm (Thousands of Persons)

2	36	5	0	USPRIV	All Employees: Total Private Industries (Thousands of Persons)
3	37	5	1	MANEMP	All Employees: Manufacturing (Thousands of Persons)
4	38	5	1	SRVPRD	All Employees: Service-Providing Industries (Thousands of Persons)
5	39	5	1	USGOOD	All Employees: Goods-Producing Industries (Thousands of Persons)
6	40	5	1	DMANEMP	All Employees: Durable goods (Thousands of Persons)
7	41	5	1	NDMANEMP	All Employees: Nondurable goods (Thousands of Persons)
8	42	5	1	USCONS	All Employees: Construction (Thousands of Persons)
9	43	5	0	USEHS	All Employees: Education & Health Services (Thousands of Persons)
10	44	5	1	USFIRE	All Employees: Financial Activities (Thousands of Persons)
11	45	5	0	USINFO	All Employees: Information Services (Thousands of Persons)
12	46	5	0	USPBS	All Employees: Professional & Business Services (Thousands of Persons)
13	47	5	0	USLAH	All Employees: Leisure & Hospitality (Thousands of Persons)
14	48	5	0	USSERV	All Employees: Other Services (Thousands of Persons)
15	49	5	0	USMINE	All Employees: Mining and logging (Thousands of Persons)
16	50	5	1	USTPU	All Employees: Trade, Transportation & Utilities (Thousands of Persons)
17	51	5	1	USGOVT	All Employees: Government (Thousands of Persons)
18	52	5	1	USTRADE	All Employees: Retail Trade (Thousands of Persons)
19	53	5	1	USWTRADE	All Employees: Wholesale Trade (Thousands of Persons)
20	54	5	0	CES9091000001	All Employees: Government: Federal (Thousands of Persons)
21	55	5	0	CES9092000001	All Employees: Government: State Government (Thousands of Persons)
22	56	5	0	CES9093000001	All Employees: Government: Local Government (Thousands of Persons)
23	57	5	1	CE16OV	Civilian Employment (Thousands of Persons)

24	58	2	0	CIVPART	Civilian Labor Force Participation Rate (Percent)
25	59	2	1	UNRATE	Civilian Unemployment Rate (Percent)
26	60	2	0	UNRATESTx	Unemployment Rate less than 27 weeks (Percent)
27	61	2	0	UNRATELTx	Unemployment Rate for more than 27 weeks (Percent)
28	62	2	0	LNS14000012	Unemployment Rate - 16 to 19 years (Percent)
29	63	2	0	LNS14000025	Unemployment Rate - 20 years and over, Men (Percent)
30	64	2	0	LNS14000026	Unemployment Rate - 20 years and over, Women (Percent)
31	65	5	1	UEMPLT5	Number of Civilians Unemployed - Less Than 5 Weeks (Thousands of Persons)
32	66	5	1	UEMP5TO14	Number of Civilians Unemployed for 5 to 14 Weeks (Thousands of Persons)
33	67	5	1	UEMP15T26	Number of Civilians Unemployed for 15 to 26 Weeks (Thousands of Persons)
34	68	5	1	UEMP27OV	Number of Civilians Unemployed for 27 Weeks and Over (Thousands of Persons)
35	69	5	0	LNS13023621	Unemployment Level - Job Losers (Thousands of Persons)
36	70	5	0	LNS13023557	Unemployment Level - Reentrants to Labor Force (Thousands of Persons)
37	71	5	0	LNS13023705	Unemployment Level - Job Leavers (Thousands of Persons)
38	72	5	0	LNS13023569	Unemployment Level - New Entrants (Thousands of Persons)
39	73	5	0	LNS12032194	Employment Level - Part-Time for Economic Reasons, All Industries (Thousands of persons)
40	74	5	0	HOABS	Business Sector: Hours of All Persons (Index 2012=100)
41	75	5	0	HOAMS	Manufacturing Sector: Hours of All Persons (Index 2012=100)
42	76	5	0	HOANBS	Nonfarm Business Sector: Hours of All Persons (Index 2012=100)
43	77	1	1	AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing (Hours)

44	78	2	0	AWHNONAG	Average Weekly Hours Of Production And Nonsupervisory Employees: Total private (Hours)
45	79	2	1	AWOTMAN	Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing (Hours)
46	80	1	1	HWIx	Help-Wanted Index
47	197	2	1	UEMPMEAN	Average (Mean) Duration of Unemployment (Weeks)
48	198	2	1	CES0600000007	Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing
49	220	2	1	HWIURATIOx	Ratio of Help Wanted/No. Unemployed
50	221	5	1	CLAIMSx	Initial Claims

**Table C.6:** Group 4: Housing

ID	Tcode	MD	FRED mnemonic	FRED Description	
1	81	5	1	HOUST	Housing Starts: Total: New Privately Owned Housing Units Started (Thousands of Units)
2	82	5	0	HOUST5F	Privately Owned Housing Starts: 5-Unit Structures or More (Thousands of Units)
3	83	5	1	PERMIT	New Private Housing Units Authorized by Building Permits (Thousands of Units)
4	84	5	1	HOUSTMW	Housing Starts in Midwest Census Region (Thousands of Units)
5	85	5	1	HOUSTNE	Housing Starts in Northeast Census Region (Thousands of Units)
6	86	5	1	HOUSTS	Housing Starts in South Census Region (Thousands of Units)
7	87	5	1	HOUSTW	Housing Starts in West Census Region (Thousands of Units)
8	179	5	0	USSTHPI	All-Transactions House Price Index for the United States (Index 1980 Q1=100)
9	180	5	0	SPCS10RSA	S&P/Case-Shiller 10-City Composite Home Price Index (Index January 2000 = 100)

10	181*	5	0	SPCS20RSA	S&P/Case-Shiller 20-City Composite Home Price Index (Index January 2000 = 100)
11	227	5	1	PERMITNE	New Private Housing Units Authorized by Building Permits in the Northeast Census Region (Thousands, SAAR)
12	228	5	1	PERMITMW	New Private Housing Units Authorized by Building Permits in the Midwest Census Region (Thousands, SAAR)
13	229	5	1	PERMITS	New Private Housing Units Authorized by Building Permits in the South Census Region (Thousands, SAAR)
14	230	5	1	PERMITW	New Private Housing Units Authorized by Building Permits in the West Census Region (Thousands, SAAR)

**Table C.7:** Group 5: Inventories, orders, and sales

ID	Tcode	MD	FRED mnemonic	FRED Description	
1	88	5	1	CMRMTSPLx	Real Manufacturing and Trade Industries Sales (Millions of Chained 2012 Dollars)
2	89	5	0	RSAFSx	Retail and Food Services Sales (Millions of Chained 2012 Dollars), deflated by Core PCE
3	90	5	1	AMDMNOx	Real Manufacturers' New Orders: Durable Goods (Millions of 2012 Dollars), deflated by Core PCE
4	91	5	1	ACOGNOx	Real Value of Manufacturers' New Orders for Consumer Goods Industries (Millions of 2012 Dollars), deflated by Core PCE
5	92	5	1	AMDMUOx	Real Value of Manufacturers' Unfilled Orders for Durable Goods Industries (Millions of 2012 Dollars), deflated by Core PCE
6	93	5	1	ANDENOx	Real Value of Manufacturers' New Orders for Capital Goods: Nondefense Capital Goods Industries (Millions of 2012 Dollars), deflated by Core PCE

7	94	5		INVCQRMTSPL	Real Manufacturing and Trade Inventories (Millions of 2012 Dollars)
8	222	5	1	BUSINVx	Total Business Inventories (Millions of Dollars)
9	223	2	1	ISRATIOx	Total Business: Inventories to Sales Ratio

**Table C.8:** Group 6: Prices

ID	Tcode	MD	FRED mnemonic	FRED Description	
1	95	6	0	PCECTPI	Personal Consumption Expenditures: Chain-type Price Index (Index 2012=100)
2	96	6	0	PCEPILFE	Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index) (Index 2012=100)
3	97	6	0	GDPCTPI	Gross Domestic Product: Chain-type Price Index (Index 2012=100)
4	98	6	0	GPDICTPI	Gross Private Domestic Investment: Chain-type Price Index (Index 2012=100)
5	99	6	0	IPDBS	Business Sector: Implicit Price Deflator (Index 2012=100)
6	100	6	0	DGDSRG3Q086SBEA	Personal consumption expenditures: Goods (chain-type price index)
7	101	6	1	DDURRG3Q086SBEA	Personal consumption expenditures: Durable goods (chain-type price index)
8	102	6	1	DSERRG3Q086SBEA	Personal consumption expenditures: Services (chain-type price index)
9	103	6	1	DNDGRG3Q086SBEA	Personal consumption expenditures: Nondurable goods (chain-type price index)
10	104	6	0	DHCERG3Q086SBEA	Personal consumption expenditures: Services: Household consumption expenditures (chain-type price index)
11	105	6	0	DMOTRG3Q086SBEA	Personal consumption expenditures: Durable goods: Motor vehicles and parts (chain-type price index)
12	106	6	0	DFDHRG3Q086SBEA	Personal consumption expenditures: Durable goods: Furnishings and durable household equipment (chain-type price index)

13	107	6	0	DREQRG3Q086SBEA	Personal consumption expenditures: Durable goods: Recreational goods and vehicles (chain-type price index)
14	108	6	0	DODGRG3Q086SBEA	Personal consumption expenditures: Durable goods: Other durable goods (chain-type price index)
15	109	6	0	DFXARG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Food and beverages purchased for off-premises consumption (chain-type price index)
16	110	6	0	DCLORG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Clothing and footwear (chain-type price index)
17	111	6	0	DGOERG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Gasoline and other energy goods (chain-type price index)
18	112	6	0	DONGRG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Other nondurable goods (chain-type price index)
19	113	6	0	DHUTRG3Q086SBEA	Personal consumption expenditures: Services: Housing and utilities (chain-type price index)
20	114	6	0	DHLCRG3Q086SBEA	Personal consumption expenditures: Services: Health care (chain-type price index)
21	115	6	0	DTRSRG3Q086SBEA	Personal consumption expenditures: Transportation services (chain-type price index)
22	116	6	0	DRCARG3Q086SBEA	Personal consumption expenditures: Recreation services (chain-type price index)
23	117	6	0	DFSARG3Q086SBEA	Personal consumption expenditures: Services: Food services and accommodations (chain-type price index)
24	118	6	0	DIFSRG3Q086SBEA	Personal consumption expenditures: Financial services and insurance (chain-type price index)
25	119	6	0	DOTSRG3Q086SBEA	Personal consumption expenditures: Other services (chain-type price index)
26	120	6	1	CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items (Index 1982-84=100)

27	121	6	0	CPILFESL	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy (Index 1982-84=100)
28	122	6	1	WPSFD49207	Producer Price Index by Commodity for Final Demand: Finished Goods (Index 1982=100)
29	123	6	0	PPIACO	Producer Price Index for All Commodities (Index 1982=100)
30	124	6	1	WPSFD49502	Producer Price Index by Commodity for Final Demand: Personal Consumption Goods (Finished Consumer Goods) (Index 1982=100)
31	125	6	0	WPSFD4111	Producer Price Index by Commodity for Finished Consumer Foods (Index 1982=100)
32	126	6	0	PPIIDC	Producer Price Index by Commodity Industrial Commodities (Index 1982=100)
33	127	6	1	WPSID61	Producer Price Index by Commodity Intermediate Materials: Supplies & Components (Index 1982=100)
34	128	5	0	WPU0531	Producer Price Index by Commodity for Fuels and Related Products and Power: Natural Gas (Index 1982=100)
35	129	5	0	WPU0561	Producer Price Index by Commodity for Fuels and Related Products and Power: Crude Petroleum (Domestic Production) (Index 1982=100)
36	130	5	1	OILPRICEx	Real Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma (2012 Dollars per Barrel), deflated by Core PCE
37	205	6	1	WPSID62	Producer Price Index: Crude Materials for Further Processing (Index 1982=100)
38	206	6	1	PPICMM	Producer Price Index: Commodities: Metals and metal products: Primary nonferrous metals (Index 1982=100)
39	207	6	1	CPIAPPSL	Consumer Price Index for All Urban Consumers: Apparel (Index 1982-84=100)
40	208	6	1	CPITRNSL	Consumer Price Index for All Urban Consumers: Transportation (Index 1982-84=100)

41	209	6	1	CPIMEDSL	Consumer Price Index for All Urban Consumers: Medical Care (Index 1982-84=100)
42	210	6	1	CUSR0000SAC	Consumer Price Index for All Urban Consumers: Commodities (Index 1982-84=100)
43	211	6	1	CUSR0000SAD	Consumer Price Index for All Urban Consumers: Durables (Index 1982-84=100)
44	212	6	1	CUSR0000SAS	Consumer Price Index for All Urban Consumers: Services (Index 1982-84=100)
45	213	6	1	CPIULFSL	Consumer Price Index for All Urban Consumers: All Items Less Food (Index 1982-84=100)
46	214	6	1	CUSR0000SA0L2	Consumer Price Index for All Urban Consumers: All items less shelter (Index 1982-84=100)
47	215	6	1	CUSR0000SA0L5	Consumer Price Index for All Urban Consumers: All items less medical care (Index 1982-84=100)
48	233	6	0	CUSR0000SEHC	CPI for All Urban Consumers: Owners' equivalent rent of residences (Index Dec 1982=100)

**Table C.9:** Group 7: Earnings and productivity

ID	Tcode	MD	FRED mnemonic	FRED Description	
1	131	5	AHETPIx	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private (2012 Dollars per Hour), deflated by Core PCE	
2	132	5	1	CES2000000008x	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Construction (2012 Dollars per Hour), deflated by Core PCE
3	133	5	1	CES3000000008x	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing (2012 Dollars per Hour), deflated by Core PCE
4	134	5	0	COMPRMS	Manufacturing Sector: Real Compensation Per Hour (Index 2012=100)

5	135	5	0	COMPRNFB	Nonfarm Business Sector: Real Compensation Per Hour (Index 2012=100)
6	136	5	0	RCPHBS	Business Sector: Real Compensation Per Hour (Index 2012=100)
7	137	5	0	OPHMFG	Manufacturing Sector: Real Output Per Hour of All Persons (Index 2012=100)
8	138	5	0	OPHNFB	Nonfarm Business Sector: Real Output Per Hour of All Persons (Index 2012=100)
9	139	5	0	OPHPBS	Business Sector: Real Output Per Hour of All Persons (Index 2012=100)
10	140	5	0	ULCBS	Business Sector: Unit Labor Cost (Index 2012=100)
11	141	5	0	ULCMFG	Manufacturing Sector: Unit Labor Cost (Index 2012=100)
12	142	5	0	ULCNFB	Nonfarm Business Sector: Unit Labor Cost (Index 2012=100)
13	143	5	0	UNLPNBS	Nonfarm Business Sector: Unit Non-labor Payments (Index 2012=100)
14	216	6	1	CES0600000008	Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing (Dollars per Hour)

**Table C.10:** Group 8: Interest rates

	ID	Tcode	MD	FRED mnemonic	FRED Description
1	144	2	1	FEDFUNDS	Effective Federal Funds Rate (Percent)
2	145	2	1	TB3MS	3-Month Treasury Bill: Secondary Market Rate (Percent)
3	146	2	1	TB6MS	6-Month Treasury Bill: Secondary Market Rate (Percent)
4	147	2	1	GS1	1-Year Treasury Constant Maturity Rate (Percent)
5	148	2	1	GS10	10-Year Treasury Constant Maturity Rate (Percent)
6	149	2	0	MORTGAGE30US	30-Year Conventional Mortgage Rate© (Percent)
7	150	2	1	AAA	Moody's Seasoned Aaa Corporate Bond Yield© (Percent)
8	151	2	1	BAA	Moody's Seasoned Baa Corporate Bond Yield© (Percent)

9	152	1	0	BAA10YM	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity (Percent)
10	153	1	0	MORTG10YRx	30-Year Conventional Mortgage Rate Relative to 10-Year Treasury Constant Maturity (Percent)
11	154	1	0	TB6M3Mx	6-Month Treasury Bill Minus 3-Month Treasury Bill, secondary market (Percent)
12	155	1	0	GS1TB3Mx	1-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)
13	156	1	0	GS10TB3Mx	10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)
14	157	1	0	CPF3MTB3Mx	3-Month Commercial Paper Minus 3-Month Treasury Bill, secondary market (Percent)
15	201	2	1	GS5	5-Year Treasury Constant Maturity Rate
16	202	1	1	TB3SMFFM	3-Month Treasury Constant Maturity Minus Federal Funds Rate
17	203	1	1	T5YFFM	5-Year Treasury Constant Maturity Minus Federal Funds Rate
18	204	1	1	AAAFFM	Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate
19	225	2	1	CP3M	3-Month AA Financial Commercial Paper Rate
20	226	1	1	COMPAPFF	3-Month Commercial Paper Minus Federal Funds Rate

**Table C.11:** Group 9: Money and credit

ID	Tcode	MD	FRED mnemonic	FRED Description	
1	158	5	0	BOGMBASEREALx	Monetary Base (Millions of 1982-84 Dollars), deflated by CPI
2	159	5	0	IMFSLx	Real Institutional Money Funds (Billions of 2012 Dollars), deflated by Core PCE
3	160	5	0	M1REAL	Real M1 Money Stock (Billions of 1982-84 Dollars), deflated by CPI

4	161	5	1	M2REAL	Real M2 Money Stock (Billions of 1982-84 Dollars), deflated by CPI
5	162	5	0	MZMREAL	Real MZM Money Stock (Billions of 1982-84 Dollars), deflated by CPI
6	163	5	1	BUSLOANSx	Real Commercial and Industrial Loans, All Commercial Banks (Billions of 2012 U.S. Dollars), deflated by Core PCE
7	164	5	0	CONSUMERx	Real Consumer Loans at All Commercial Banks (Billions of 2012 U.S. Dollars), deflated by Core PCE
8	165	5	1	NONREVSLx	Total Real Nonrevolving Credit Owned and Securitized, Outstanding (Billions of 2012 Dollars), deflated by Core PCE
9	166	5	1	REALLNx	Real Real Estate Loans, All Commercial Banks (Billions of 2012 U.S. Dollars), deflated by Core PCE
10	167	5	0	REVOLSLx	Total Real Revolving Credit Owned and Securitized, Outstanding (Billions of 2012 Dollars), deflated by Core PCE
11	168	5	0	TOTALSLx	Total Consumer Credit Outstanding (Billions of 2012 Dollars), deflated by Core PCE
12	169	1	0	DRIWCIL	FRB Senior Loans Officer Opions. Net Percentage of Domestic Respondents Reporting Increased Willingness to Make Consumer Installment Loans
13	199	6	1	TOTRESNS	Total Reserves of Depository Institutions (Billions of Dollars)
14	200	7	1	NONBORRES	Reserves Of Depository Institutions, Nonborrowed (Millions of Dollars)
15	217	6	1	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding Owned by Finance Companies (Millions of Dollars)
16	218	6	1	DTCTHFNM	Total Consumer Loans and Leases Outstanding Owned and Securitized by Finance Companies (Millions of Dollars)
17	219	6	1	INVEST	Securities in Bank Credit at All Commercial Banks (Billions of Dollars)

**Table C.12:** Group 10: Household balance sheets

	ID	Tcode	MD	FRED mnemonic	FRED Description
1	170	5	0	TABSHNOx	Real Total Assets of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE
2	171	5	0	TLBSHNOx	Real Total Liabilities of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE
3	172	5	0	LIABPIx	Liabilities of Households and Nonprofit Organizations Relative to Personal Disposable Income (Percent)
4	173	5	0	TNWBSHNOx	Real Net Worth of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE
5	174	1	0	NWPIx	Net Worth of Households and Nonprofit Organizations Relative to Disposable Personal Income (Percent)
6	175	5	0	TARESAx	Real Assets of Households and Nonprofit Organizations excluding Real Estate Assets (Billions of 2012 Dollars), deflated by Core PCE
7	176	5	0	HNOREMQ027Sx	Real Real Estate Assets of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE
8	177	5	0	TFAABSHNOx	Real Total Financial Assets of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE
9	224	2	1	CONSPIx	Nonrevolving consumer credit to Personal Income

**Table C.13:** Group 11: Exchange rates

	ID	Tcode	MD	FRED mnemonic	FRED Description
1	182	5	1	TWEXAFEGSMTHx	Trade Weighted U.S. Dollar Index: Advanced Foreign Currencies (Index Jan 2006=100)
2	183*	5	0	EXUSEU	U.S. / Euro Foreign Exchange Rate (U.S. Dollars to One Euro)
3	184	5	1	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate
4	185	5	1	EXJPUSx	Japan / U.S. Foreign Exchange Rate

5	186	5	1	EXUSUKx	U.S. / U.K. Foreign Exchange Rate
6	187	5	1	EXCAUSx	Canada / U.S. Foreign Exchange Rate

**Table C.14:** Group 12: Other

	ID	Tcode	MD	FRED mnemonic	FRED Description
1	188	1	1	UMCSENTx	University of Michigan: Consumer Sentiment (Index 1st Quarter 1966=100)
2	189	2	0	USEPUINDXM	Economic Policy Uncertainty Index for United States

**Table C.15:** Group 13: Stock markets

	ID	Tcode	MD	FRED mnemonic	FRED Description
1	178	1	1	VXOCLSX	CBOE S&P 100 Volatility Index: V XO
2	231	5	0	NIKKEI225	Nikkei Stock Average
3	232	5	0	NASDAQCOM	NASDAQ Composite (Index Feb 5, 1971=100)
4	245	5	1	S&P 500	S&P's Common Stock Price Index: Composite
5	246	5	1	S&P: indust	S&P's Common Stock Price Index: Industrials
6	247	2	1	S&P: div yield	S&P's Composite Common Stock: Dividend Yield
7	248	5	1	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio

**Table C.16:** Group 14: Non-household balance sheets

	ID	Tcode	MD	FRED mnemonic	FRED Description
1	192	2	0	GFDEGDQ188S	Federal Debt: Total Public Debt as Percent of GDP (Percent)
2	193	2	0	GFDEBTNx	Real Federal Debt: Total Public Debt (Millions of 2012 Dollars), deflated by PCE

3	234	5	0	TLBSNNCBx	Real Nonfinancial Corporate Business Sector Liabilities (Billions of 2012 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
4	235	1	0	TLBSNNCBBDIx	Nonfinancial Corporate Business Sector Liabilities to Disposable Business Income (Percent)
5	236	5	0	TTAABSNNCBx	Real Nonfinancial Corporate Business Sector Assets (Billions of 2012 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
6	237	5	0	TNWMVBSNNCBx	Real Nonfinancial Corporate Business Sector Net Worth (Billions of 2012 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
7	238	2	0	TNWMVBSNNCBBDIx	Nonfinancial Corporate Business Sector Net Worth to Disposable Business Income (Percent)
8	239	5	0	TLBSNNBx Real	Nonfinancial Noncorporate Business Sector Liabilities (Billions of 2012 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
9	240	1	0	TLBSNNBBDIx	Nonfinancial Noncorporate Business Sector Liabilities to Disposable Business Income (Percent)
10	241	5	0	TABSNNBx	Real Nonfinancial Noncorporate Business Sector Assets (Billions of 2012 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
11	242	5	0	TNWBSNNBx	Real Nonfinancial Noncorporate Business Sector Net Worth (Billions of 2012 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS
12	243	2	0	TNWBSNNBBDIx	Nonfinancial Noncorporate Business Sector Net Worth to Disposable Business Income (Percent)
13	244	5	0	CNCFx	Real Disposable Business Income, Billions of 2012 Dollars (Corporate cash flow With IVA minus taxes on corporate income, deflated by Implicit Price Deflator for Business Sector IPDBS)

## C.4 Monthly data description

Data on monthly macroeconomic variables is from the "FRED-MD" database, put together by McCracken and Ng (2016). The paper relies on data from the FRED-MD vintage as of April, 2021. Following McCracken and Ng (2016), the data in FRED-MD can be grouped into 8 distinct groups, provided in the tables below, which are adapted from McCracken and Ng (2016): "column ID" denotes series number; "Tcode" denotes one of the following stationarity transformations: (1) no transformation; (2)  $\Delta x_{t_m}$ ; (3)  $\Delta^2 x_{t_m}$ ; (4)  $\log x_{t_m}$ ; (5)  $\Delta \log x_{t_m}$ ; (6)  $\Delta^2 \log x_{t_m}$ ; (7)  $\Delta(x_{t_m} / x_{t_m-1} - 1.0)$ ; "FRED mnemonic" is mnemonics used by the Federal Reserve Bank of St. Louis; "FRED are descriptions" are descriptions used by the Federal Reserve Bank of St. Louis. An asterisk on ID-number is used to indicate series with no observations available for estimation in the initial pseudo vintage, as discussed in the main text.

**Table C.17:** Group 1: Output and income

	ID	Tcode	FRED mnemonic	FRED Description
1	1	5	RPI	Real Personal Income
2	2	5	W875RX1	Real personal income ex transfer receipts
3	6	5	INDPRO	IP Index
4	7	5	IPFPNSS	IP: Final Products and Nonindustrial Supplies
5	8	5	IPFINAL	IP: Final Products (Market Group)
6	9	5	IPCONGD	IP: Consumer Goods
7	10	5	IPDCONGD	IP: Durable Consumer Goods
8	11	5	IPNCONGD	IP: Nondurable Consumer Goods
9	12	5	IPBUSEQ	IP: Business Equipment
10	13	5	IPMAT	IP: Materials
11	14	5	IPDMAT	IP: Durable Materials
12	15	5	IPNMAT	IP: Nondurable Materials
13	16	5	IPMANSICS	IP: Manufacturing (SIC)
14	17	5	IPB51222s	IP: Residential Utilities
15	18	5	IPFUELS	IP: Fuels
16	20	2	CUMFNS	Capacity Utilization: Manufacturing

**Table C.18:** Group 2: Labor market

	ID	Tcode	FRED mnemonic	FRED Description
1	21	2	HWI	Help-Wanted Index for United States
2	22	2	HWIURATIO	Ratio of Help Wanted/No. Unemployed

3	23	5	CLF16OV	Civilian Labor Force
4	24	5	CE16OV	Civilian Employment
5	25	2	UNRATE	Civilian Unemployment Rate
6	26	2	UEMPMEAN	Average Duration of Unemployment (Weeks)
7	27	5	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks
8	28	5	UEMP5TO14	Civilians Unemployed for 5-14 Weeks
9	29	5	UEMP15OV	Civilians Unemployed - 15 Weeks & Over
10	30	5	UEMP15T26	Civilians Unemployed for 15-26 Weeks
11	31	5	UEMP27OV	Civilians Unemployed for 27 Weeks and Over
12	32	5	CLAIMSX	Initial Claims
13	33	5	PAYEMS	All Employees: Total nonfarm
14	34	5	USGOOD	All Employees: Goods-Producing Industries
15	35	5	CES1021000001	All Employees: Mining and Logging: Mining
16	36	5	USCONS	All Employees: Construction
17	37	5	MANEMP	All Employees: Manufacturing
18	38	5	DMANEMP	All Employees: Durable goods
19	39	5	NDMANEMP	All Employees: Nondurable goods
20	40	5	SRVPRD	All Employees: Service-Providing Industries
21	41	5	USTPU	All Employees: Trade, Transportation & Utilities
22	42	5	USWTRADE	All Employees: Wholesale Trade
23	43	5	USTRADE	All Employees: Retail Trade
24	44	5	USFIRE	All Employees: Financial Activities
25	45	5	USGOVT	All Employees: Government
26	46	1	CES0600000007	Avg Weekly Hours : Goods-Producing
27	47	2	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing
28	48	1	AWHMAN	Avg Weekly Hours : Manufacturing
29	127	6	CES0600000008	Avg Hourly Earnings : Goods-Producing
30	128	6	CES2000000008	Avg Hourly Earnings : Construction
31	129	6	CES3000000008	Avg Hourly Earnings : Manufacturing

**Table C.19:** Group 3: Housing

ID	Tcode	FRED mnemonic	FRED Description
----	-------	---------------	------------------

1	50	4	HOUST	Housing Starts: Total New Privately Owned
2	51	4	HOUSTNE	Housing Starts, Northeast
3	52	4	HOUSTMW	Housing Starts, Midwest
4	53	4	HOUSTS	Housing Starts, South
5	54	4	HOUSTW	Housing Starts, West
6	55	4	PERMIT	New Private Housing Permits (SAAR)
7	56	4	PERMITNE	New Private Housing Permits, Northeast (SAAR)
8	57	4	PERMITMW	New Private Housing Permits, Midwest (SAAR)
9	58	4	PERMITS	New Private Housing Permits, South (SAAR)
10	59	4	PERMITW	New Private Housing Permits, West (SAAR)

**Table C.20:** Group 4: Consumption, orders, and inventories

ID	Tcode	FRED mnemonic	FRED Description
1	3	5	DPCERA3M086SBEA
			Real personal consumption expenditures
2	4	5	CMRMTSPLx
3	5	5	RETAILx
4	64*	5	ACOGNO
5	65	5	AMDMNOx
6	66	5	ANDENOx
			New Orders for Consumer Goods
			New Orders for Durable Goods
			New Orders for Nondefense Capital Goods
7	67	5	AMDMUOx
8	68	5	BUSINVx
9	69	2	ISRATIOx
10	130	2	UMCSENTx
			Total Business Inventories
			Total Business: Inventories to Sales Ratio
			Consumer Sentiment Index

**Table C.21:** Group 5: Money and credit

ID	Tcode	FRED mnemonic	FRED Description
1	70	6	M1SL
2	71	6	M2SL
3	72	5	M2REAL
4	73	6	BOGMBASE
			M1 Money Stock
			M2 Money Stock
			Real M2 Money Stock
			Monetary Base

5	74	6	TOTRESNS	Total Reserves of Depository Institutions
6	75	7	NONBORRES	Reserves Of Depository Institutions
7	76	6	BUSLOANS	Commercial and Industrial Loans
8	77	6	REALLN	Real Estate Loans at All Commercial Banks
9	78	6	NONREVSL	Total Nonrevolving Credit
10	79	2	CONSPI	Nonrevolving consumer credit to Personal Income
11	131	6	MZMSL	MZM Money Stock
12	132	6	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding
13	133	6	DTCTHFNM	Total Consumer Loans and Leases Outstanding
14	134	6	INVEST	Securities in Bank Credit at All Commercial Banks

**Table C.22:** Group 6: Interest and exchange rates

	ID	Tcode	FRED mnemonic	FRED Description
1	84	2	FEDFUNDS	Effective Federal Funds Rate
2	85	2	CP3Mx	3-Month AA Financial Commercial Paper Rate
3	86	2	TB3MS	3-Month Treasury Bill
4	87	2	TB6MS	6-Month Treasury Bill
5	88	2	GS1	1-Year Treasury Rate
6	89	2	GS5	5-Year Treasury Rate
7	90	2	GS10	10-Year Treasury Rate
8	91	2	AAA	Moody's Seasoned Aaa Corporate Bond Yield
9	92	2	BAA	Moody's Seasoned Baa Corporate Bond Yield
10	93	1	COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS
11	94	1	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS
12	95	1	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS
13	96	1	T1YFFM	1-Year Treasury C Minus FEDFUNDS
14	97	1	T5YFFM	5-Year Treasury C Minus FEDFUNDS
15	98	1	T10YFFM	10-Year Treasury C Minus FEDFUNDS
16	99	1	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS
17	100	1	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS

18	101	5	TWEXAFEGSMTHx	Trade Weighted U.S. Dollar Index
19	102	5	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate
20	103	5	EXJPUSx	Japan / U.S. Foreign Exchange Rate
21	104	5	EXUSUKx	U.S. / U.K. Foreign Exchange Rate
22	105	5	EXCAUSx	Canada / U.S. Foreign Exchange Rate

**Table C.23:** Group 7: Prices

ID	Tcode	FRED mnemonic	FRED Description
1	106	6	WPSFD49207
2	107	6	WPSFD49502
3	108	6	WPSID61
4	109	6	WPSID62
5	110*	6	OILPRICEEx
6	111	6	PPICMM
7	113	6	CPIAUCSL
8	114	6	CPIAPPSL
9	115	6	CPITRNSL
10	116	6	CPIMEDSL
11	117	6	CUSR0000SAC
12	118	6	CUSR0000SAD
13	119	6	CUSR0000SAS
14	120	6	CPIULFSL
15	121	6	CUSR0000SA0L2
16	122	6	CUSR0000SA0L5
17	123	6	PCEPI
18	124	6	DDURRG3M086SBEA
19	125	6	DNDGRG3M086SBEA
20	126	6	DSERRG3M086SBEA

**Table C.24:** Group 8: Stock market

ID	Tcode	FRED mnemonic	FRED Description
1	80	5	S&P 500
2	81	5	S&P: indust

3	82	2	S&P div yield	S&P's Composite Common Stock: Dividend Yield
4	83	5	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio
5	135	1	VXOCLSX	VXO

---

---

## Declaration of co-authorship\*

Full name of the PhD student: Sebastian Mathias Jensen

This declaration concerns the following article/manuscript:

Title:	A Neural Network Approach to the Environmental Kuznets Curve
Authors:	Mikkel Bennedsen, Eric Hillebrand, and Sebastian Mathias Jensen

The article/manuscript is: Published  Accepted  Submitted  In preparation

If published, state full reference:

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No  Yes  If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

- A. Has essentially done all the work
- B. Major contribution
- C. Equal contribution
- D. Minor contribution
- E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	B
3. Involvement in the experimental work/clinical studies/data collection	A
4. Interpretation of the results	B
5. Writing of the first draft of the manuscript	A
6. Finalization of the manuscript and submission	B

### Signatures of the co-authors

Date	Name	Signature
20.08.21	Eric Hillebrand	
20.08.21	Mikkel Bennedsen	

In case of further co-authors please attach appendix

Date: 31.08.21



Signature of the PhD student

\*As per policy the co-author statement will be published with the dissertation.

## Declaration of co-authorship\*

Full name of the PhD student: Sebastian Mathias Jensen

This declaration concerns the following article/manuscript:

Title:	Apocalypse Now? Projecting CO <sub>2</sub> Emissions with Neural Networks
Authors:	Mikkel Bennedsen, Eric Hillebrand, and Sebastian Mathias Jensen

The article/manuscript is: Published  Accepted  Submitted  In preparation

If published, state full reference:

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No  Yes  If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

- A. Has essentially done all the work
- B. Major contribution
- C. Equal contribution
- D. Minor contribution
- E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	B
3. Involvement in the experimental work/clinical studies/data collection	A
4. Interpretation of the results	B
5. Writing of the first draft of the manuscript	A
6. Finalization of the manuscript and submission	B

### Signatures of the co-authors

Date	Name	Signature
20.08.21	Eric Hillebrand	
20.08.21	Mikkel Bennedsen	

In case of further co-authors please attach appendix

Date: 31.08.21



Signature of the PhD student

\*As per policy the co-author statement will be published with the dissertation.