Antony Millner and Thomas K. J. McDermott

Model confirmation in climate economics

Article (Accepted version) (Refereed)

Original citation: Millner, Antony and McDermott, Thomas K. J. (2016) *Model confirmation in climate economics.* Proceedings of the National Academy of Sciences of the United States of America, 113 (31). pp. 8675-8680. ISSN 1091-6490

DOI: 10.1073/pnas.1604121113

On Model Confirmation in Climate Economics

Antony Millner *, **Thomas K.J. McDermott** †

*Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, Houghton St, London, WC2A 2AE, UK United Kingdom, and †School of Economics, University College Cork, Cork, Ireland

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Benefit-cost Integrated assessment models (BC-IAMs) inform climate policy debates by quantifying the tradeoffs between alternative greenhouse gas abatement options. They achieve this by coupling simplified models of the climate system to models of the global economy and the costs and benefits of climate policy. While these models have provided valuable qualitative insights into the sensitivity of policy tradeoffs to different ethical and empirical assumptions, they are increasingly being used to inform the selection of policies in the real world. To the extent that BC-IAMs are used as inputs to policy selection, our confidence in their quantitative outputs must depend on the empirical validity of their modeling assumptions. We have a degree of confidence in climate models both because they have been tested on historical data in hindcasting experiments, and because the physical principles they are based on have been empirically confirmed in closely related applications. By contrast, the economic components of BC-IAMs often rely on untestable scenarios, or on structural models that are comparatively untested on relevant time scales. Where possible, an approach to model confirmation similar to that used in climate science could help to build confidence in the economic components of BC-IAMs, or focus attention on which components might need refinement for policy applications. We illustrate the potential benefits of model confirmation exercises by performing a long-run hindcasting experiment with one of the leading BC-IAMs. We show that its model of long-run economic growth – one of its most important economic components – had questionable predictive power over the 20th century.

Integrated assessment | climate policy | long-run economic growth | model confirmation | structural uncertainty

Abbreviations: IAM, integrated assessment model; BC-IAM, benefit-cost integrated assessment model; SCC, social cost of carbon; TFP, total factor productivity

'Prediction is very difficult, especially about the future.' – Niels Bohr

A little over twenty years ago a seminal article on the interpretation of numerical models in the earth sciences appeared in a leading scientific journal [1]. The authors argued that while verification and validation of these models is strictly logically impossible, model confirmation is a necessary and desirable step. In the intervening years an impressive body of work in climate science has compared the predictions of global climate models with observations. Chapter 9 of the Intergovermental Panel on Climate Change's Fifth Assessment report summarizes recent work [2], stating that "model evaluation...reflects the need for climate models to represent the observed behaviour of past climate as a necessary condition to be considered a viable tool for future projections". Scientists continue to use empirical tests of climate models to refine and improve them, while also reflecting on the methodological questions that arise when interpreting model predictions to inform decision-making [3].

Climate models are however only a part of the technical apparatus that has been developed to inform climate policy decisions. Integrated Assessment Models (IAMs) provide the link between physical science and policy. IAMs come in two varieties – benefit-cost models, which attempt to estimate the aggregate costs and benefits of climate policy to society, and detailed process models, which usually analyze more detailed policies in a cost-effectiveness framework (i.e. assuming an exogenous policy objective), often in much greater sectoral detail than the highly aggregated benefit-cost models [4]. We focus on Benefit-Cost IAMs (BC-IAMs) in this article, as these have been the focus of research activity in economics [5, 6], and are increasingly influential in policy applications.

BC-IAMs couple simplified climate models with representations of the global economy in an attempt to understand the tradeoffs between alternative policy options. They have been applied to a wide variety of questions: How might different welfare frameworks affect the attractiveness of policy options (e.g. [7])? Which approaches to international agreements are likely to succeed (e.g. [8])? How might different policy instruments affect innovation in energy technologies (e.g. [9])? These modeling exercises provide valuable insights into the possible qualitative differences between policy options. However, their *quantitative* implications are conditional on the veracity of the underlying models. BC-IAMs can be used to show that policy A leads to higher welfare than policy B in model X, but in order to extend this model-based finding to claims about reality, we need to know how well X approximates reality. Are the equations and initialization procedures used by model X *structurally sound*, and if not, what risks might we run by treating them as if they are?[1]

Significance

Benefit-Cost Integrated assessment models (BC-IAMs) combine climate science, impacts studies, and representations of long-run economic growth to estimate the costs and benefits of climate policy. They provide valuable qualitative insights into how policy outcomes might depend on ethical and empirical assumptions. Increasingly however, BC-IAMs are being used to inform quantitative policy choices. Yet, their economic components are largely untested, or untestable, over relevant time scales. We demonstrate the potential benefits of model confirmation exercises for policy applications, demonstrating that the economic growth model used by a prominent BC-IAM had little predictive power over the 20th century. Insofar as possible, out-of-sample empirical tests of the economic components of BC-IAMs should inform their future development for real world applications.
Author contributions: AM and TM developed the concept of the study and analyzed the data. AM wrote the paper.

Reserved for Publication Footnotes

[1] The concept of a structurally sound BC-IAM needs to be interpreted with care. We give a detailed definition that makes our usage of this term more precise in the Supporting Information. In particular, our definition separates the empirical relationships in a BC-IAM (the focus of this article) from its ethical assumptions.

Assessing the structural soundness of economic modeling assumptions in BC-IAMs has recently become an increasingly pressing issue, as they are beginning to be used to inform quantitative real world policy decisions. For example, the US government has recently established an interagency working group [10] to estimate a value of the social cost of carbon (SCC), the welfare cost to society from emitting a ton of CO_2. The value of the SCC that was adopted will form part of the cost-benefit assessment of all federal projects and policies, and thus has the potential to influence billions of dollars of investment. The process used to establish a value for the SCC relied heavily on BC-IAMs, the first time they have directly informed quantitative federal rules. While the US SCC estimate is perhaps the most prominent recent example, other governments and international organizations are also increasingly turning to BC-IAMs to inform policy choices.

As soon as a model is used to inform quantitative policy decisions, the criteria by which it must be judged become more demanding. A given model may be a useful tool for exploring the qualitative implications of different assumptions, but in order for it to be profitably applied to policy choices, we need to know how plausible those assumptions are as empirical hypotheses about how the world works. If a model can be shown to be structurally flawed in hindcasting exercises, our expectation should be that similar errors might occur when using it to make predictions that inform policy choices today. No model is perfect, and we should not expect any given model to insure us against regret entirely. But to the extent possible, it is in our interests to attempt to ascertain how wrong we might go when relying on a given model to make decisions. As has occurred in climate science, this exercise could build confidence in those economic modeling assumptions that are found to be consistent with empirical data, and focus attention on those assumptions that may require refinement for policy applications.

Importantly, confirmation exercises provide entirely different information about a model's validity from model calibration, sensitivity analysis, probabilistic approaches to quantifying parametric uncertainty, or expert elicitation of model parameters, all of which are standard practice in the field. These uncertainty quantification methods explore the space of model outcomes (and perhaps estimate their likelihood), taking the model's structural assumptions as given. Model confirmation, on the other hand, tests whether the equations and initialization procedures a model uses to generate predictions are able to provide a good representation of observed outcomes. A model whose outcome space has been explored using the uncertainty quantification methods mentioned above may still yield error-prone predictions if the underlying modeling assumptions are not a good fit to reality. While these methods can of course generate a distribution of model outcomes, whether or not such distributions reflect the uncertainty we actually face depends on the structural soundness of the model used to generate them. We note that model confirmation is only possible if a model is specified in a self-contained manner, i.e. it is comprised of a set of structural assumptions and free parameters that can, at least in principle, be estimated from data. This makes models that rely on fixed external scenarios for generating predictions very difficult to confirm *ex ante*. Although such models could provide a good characterization of current uncertainty, we have no way of assessing whether this is likely to be the case by testing their past performance.

While the physical science models upon which the scientific components of BC-IAMs are based have often been subjected to tests of structural validity, their economic components are often either based on untestable exogenous scenarios, or on structural modeling assumptions that are largely untested on the temporal scales that are relevant to climate applications. In part this reflects genuine data difficulties, which make some economic assumptions in BC-IAMs very difficult to confirm. For example, BC-IAMs assume a functional form for the climate damage function, which quantifies the impact of global average temperature changes on the aggregate productivity of the economy. BC-IAM results are highly sensitive to the rate of increase of damages with temperature at high temperatures, but as we've only seen a small amount of average warming so far, it is very difficult to test any assumed functional form for damages. Some of the most important economic components of some BC-IAMs are however amenable to empirical tests.

An out-of-sample test of a model of long-run economic growth

To demonstrate what may be learned from model confirmation exercises we focus on the economic growth model used by the well known DICE BC-IAM [15]. The assumptions BC-IAMs make about long-run economic growth have a very substantial effect on leading policy outputs such as the SCC. This is because economic growth strongly affects the path of GHG emissions, the magnitude of climate damages, and the wealth of future generations, all key determinants of the aggregate costs and benefits of climate policy. Unlike other well known BC-IAMs (e.g. the PAGE [16] and FUND [17] models), which rely on external scenarios for economic growth that are impossible to test empirically *ex ante*, DICE uses an explicit model of economic growth that makes it well suited to empirical testing, and is also widely deployed across climate economics (see e.g. [18]). A crucial part of this growth model is a model of the temporal evolution of total factor productivity (TFP), a quantity that sets the overall level of technological advancement in the economy. Economic growth is largely driven by technological progress in DICE. Thus, although policy evaluation in DICE is also highly sensitive to other structural modeling assumptions (e.g. the shape of the damage function, and the evolution of abatement costs), a lot depends on how it models overall technological progress.[2]

In order to test the structural assumptions and initialization procedures employed by DICE's economic growth model, we consider the following question: How would this model fare if we asked it to predict the growth path of a major economy over the 20th century? This question is closely analogous to those asked of climate models by climate scientists [2]. The model of the evolution of the economy DICE employs is a version of the Ramsey neoclassical growth model, familiar to any student of macroeconomics (see [19] for a detailed exposition). In this model economic output is generated by competitive firms, and

[2] This view is confirmed by Nordhaus [12]: 'the major factor producing different climate outcomes in our uncertainty runs is differential technological change. In our estimates, the productivity uncertainty outweighs the uncertainties of the climate system and the damage function in determining the relationship between temperature change and consumption.' A global sensitivity analysis of the DICE model confirms that its SCC estimates are highly sensitive to the growth rate of TFP [13]. A heuristic understanding of why policy recommendations are so sensitive to assumptions about TFP growth can be obtained by studying the social discount rate $\rho(t)$. Under standard assumptions the change in social welfare that arises from a small change in consumption Δ_t which occurs t years in the future is given by $\Delta_t e^{-\rho(t)t}$. Standard computations (see e.g. [14]) show that in a deterministic setting $\rho(t) = \delta + \eta g(t)$, where δ is the pure rate of social time preference, η is the elasticity of marginal utility, and $g(t)$ is the average consumption growth rate between the present and time t. In most cases the term $\eta g(t)$ is the dominant contribution to $\rho(t)$. Since consumption growth $g(t)$ is driven by TFP growth in DICE, the present value of future climate damages is highly sensitive to TFP growth. For example, for $\delta = 1\%$/yr, $\eta = 2$ and $g(100) = 1\%$/yr an incremental climate damage of \$100 that occurs 100 years from now will be valued at $\$100e^{-(0.01+2\times0.01)100} \approx \$5$ in present value terms. For $g(100) = 2\%$/yr however, the same \$1 damage would be worth $\$100e^{-(0.01+2\times0.02)100} \approx \$0.7$. Thus an increase in consumption growth from 1% to 2%/yr reduces the current welfare cost of climate damages that occur in 100 years by a factor larger than 7.

is either consumed, or reinvested in firms. Firms produce output via a production technology, which uses the capital and labour supplied by consumers as inputs. In DICE technological progress is modeled as an increasing trend in total factor productivity, which acts as a multiplier on firms' production technologies. Thus as TFP grows, and the technologies of production become more advanced, fewer capital and labour inputs are required to generate a given level of economic output. A specific model of the time dependence of TFP is assumed in DICE. This model depends on free parameters that can be estimated from economic history.[3]

We test this model's predictive performance using recently compiled data on the US economy from 1870-2010 [21]. We single out the US as it is the largest economy for which detailed long-run economic data are available, and because of its position at the technological frontier over much of the 20th century [22]. Our tests are as generous as possible to the model (e.g. we assume a perfect forecast of labour supply), and stick closely to the calibration and forecasting methodology utilized by DICE (details of the model implementation are available in the Supporting Information). Due to the generosity of our modeling assumptions our results likely exaggerate the model's predictive performance. To test the model's predictive power we divide the data into different training and verification windows. 95% confidence intervals (CI) for the parameters of the TFP model are inferred from the training data. The state equation for the capital stock and empirically estimated model of TFP evolution are then used to predict economic output.

Fig. 1 is illustrative of our model estimation and confirmation methodology. The figure depicts a long-run forecast of TFP and economic output obtained by estimating the TFP model on the 50 years of data from 1870 to 1920. The left panel of the figure depicts the fit of the TFP model to the training data, and its out-of-sample projection of TFP. The right panel shows an out-of-sample projection of GDP at 1920, which is generated using the empirically estimated TFP model, the state equation for the evolution of the capital stock, and a perfect forecast of labour supply. The figure shows that although the TFP model fits the training data well, its out-of-sample forecast substantially underestimates technological progress in the latter half of the 20th century. These errors are compounded for GDP projections, as persistent underestimates of TFP affect predicted investment flows and capital formation in each future period, which further downward bias the model. We note however that while the presence of model errors is significant, the fact that the model was downward biased in 1920 does not imply that it will be downward biased in all periods, as we demonstrate below.

While the out-of-sample forecasts of TFP and GDP the model generates in 1920 are not successful, they nevertheless look reasonable when viewed from the perspective of the fifty year data series up to that year. Had economists produced these predictions at the time based on only these 50 years of data, they would no doubt have been perceived as plausible future growth scenarios. From today's perspective however, the model looks like a less reliable predictive tool. An important reason why the model performs poorly is that the post World War Two boost in productivity growth is not presaged in the training data at 1920. This finding is indicative of the difficulty of predicting long-run technological developments. We face precisely the same difficulties today when using BC-IAMs to project economic growth into the next century and beyond (see Supporting Information for further discussion).

While Figure 1 suggests that the model's long-run predictive performance could be a concern, it focusses on only a single forecasting date, i.e. 1920. In addition, even if the model's long-run predictions are flawed, it could be a useful predictive tool on intermediate time scales, e.g. 30 years, where unpredictable technological jumps are less likely to make past data unrepresentative of future outcomes. Although 30 years may seem a short forecast horizon for a problem as long-lived as climate change, fully 50% of the value of the SCC is determined by outcomes over this period under common parameterizations of the DICE model [23].[4] To address these issues, Fig. 2 extends the analysis of Fig. 1, summarizing the model's predictive performance at each year in the data series, for 30 and 50 year training and confirmation windows. For each year in the period 1900-1980 (1920-1960) the model was trained on the previous 30 (50) years of data, and the estimated model used to forecast GDP for the next 30 (50) years. Although the model performs well in some 30 year periods, in most years the realized growth outcome falls outside of the forecasted interval. Arguably, the model is thus not a successful predictive tool on this shorter time scale, despite less sensitivity to large unpredictable shifts in the technological frontier. For 50 year forecasts, the model performs well post World War II, but poorly in the pre-war period. This shows that the illustration of model performance depicted in Fig. 1 is not exceptional. The 50 year forecasts further illustrate the sensitivity of growth projections to structural breaks of the kind that followed the war, and demonstrate the value of a long historical perspective. Had we only evaluated the model on the most recent 60 years of data we would likely have overestimated its long-run predictive performance.

Our analysis suggests that the version of the neoclassical growth model that DICE relies on could be subject to structural errors on the temporal scales relevant to climate polices. The Ramsey growth model, and more complex models that endogenize the process of technical change, have been profitably applied to a variety of empirical questions in macroeconomics. It is thus important to understand how the use of these models in DICE and other climate applications differs from their more standard empirical applications. Growth models are usually used in empirical applications to *explain cross-country differences* between the historical growth paths of different countries. In BC-IAMs these models are used to *predict the absolute level* of global or regional economic output over the coming centuries. When neoclassical growth models are used to explain differences in past outcomes across countries, technical change, in the form of the growth rate of TFP, is formally nothing more than a residual in a linear regression. It is the part of empirical growth data that is *not* explained by the endogenous factors in the model, i.e. the productivity of capital and labour. If however such models are used to make predictions, as in DICE, the future realizations of TFP must also be predicted. This requires us to posit an explicit quantitative model of the evolution of TFP over the coming decades. Yet we have no law-like theory of long-run technical change that parallels the predictive successes that have been achieved in the physical sciences [24]. This seems unlikely to change in the near future, and more sophisticated models that endogenize the process of technological change also seem unlikely to provide high-powered predictive tools, despite their

[3] DICE assumes that the growth rate of TFP is an exponential function that slowly decays from an initial value to a smaller long-run value. The free parameters are the initial growth rate and the rate of decline of the growth rate. A large literature has developed endogenous growth models, which relate the evolution of TFP to endogenous economic variables (see [19], and [20] for a review of applications to climate economics). Since we wish to stay as close as possible to the methodology employed by DICE, we do not investigate the empirical performance of these models here, but see comments below and Section 5 in the Supporting Information.

[4] This finding is dependent on choices of welfare parameters, which in turn affect the social discount rate. All else equal, lower (higher) social discount rates make SCC values more (less) dependent on forecasts of the near future.

more nuanced representation of its causes.[5] Just as natural selection explains differences in species' phenotypes without predicting future adaptations, so growth theory has proven to be an insightful tool for explaining the causal determinants of cross-country difference in historical growth outcomes. Prediction, however, is a different matter.

Implications for the development and use of BC-IAMs in quantitative policy applications

What can be concluded from this first example of a model confirmation exercise for the economic components of a BC-IAM? Of course, DICE is a single (albeit prominent) example of a BC-IAM, and its implementation of the Ramsey model a single (albeit frequently deployed) representation of the process of long-run economic growth. Our findings are not necessarily representative of how other growth models might fare in similar confirmation exercises, the structural validity of other BC-IAMs, or the performance of their economic components. Our point however is that because most of the economic components of BC-IAMs have not, to our knowledge, been subjected to empirical tests of structural validity using historical data series, we do not know what their empirical status is for quantitative policy applications. Without testing models in hindcasting tasks closely related to the uses we wish to put them to today, we cannot gauge the extent of any possible model errors.

We close with four recommendations for the testing and use of BC-IAMs in policy applications. First, those components of BC-IAMs whose structural properties can be meaningfully tested using historical data should be. Confirmation exercises can build confidence in model components that perform well historically, and indicate the range of model parameters that needs to be considered for a given model to have a chance of making sensible out-of-sample forecasts. If no such parameter range can be identified from such an exercise, the model's structural assumptions might need to be revised for the purposes of policy applications. Calibration (a within sample exercise) and parametric uncertainty quantification techniques are not substitutes for this procedure, as we need to test the out-of-sample performance of a model's structural assumptions. This is not to say however that untested or structurally flawed models cannot be useful for illustrating qualitative conceptual points about alternative modeling assumptions. Many BC-IAMs are used profitably for this purpose in the academic community today, and none of our points undermines the value of such modeling exercises if they are interpreted with sufficient carefulness. The hurdles a model must jump over to qualify for application to quantitative policy choices should however be more demanding.

Second, the economic components of BC-IAMs should be based on testable structural hypotheses, in so far as this is possible. The confirmation exercise we conducted with the DICE model was only possible because the model is specified in a self-contained manner, and with sufficient structural detail to allow it to be meaningfully compared to data. This is a great virtue of the DICE model. While in this case our confirmation exercise raised the possibility of quantitatively meaningful errors when applying this model to policy questions, we were at least able to ask (and partially answer) the question: what risks might we run by assuming that the world behaves in line with the model? Models that rely more heavily on external scenarios for key economic components do not allow for this kind of empirical testing. If a model component cannot be tested, we cannot hope to gain confidence in it *ex ante*, even if it in fact turns out to perform well *ex post*. Thus, although a set of exogenous scenarios could turn out to capture our underlying uncertainty, we can never estimate what risks we might run by assuming this to be the case when making decisions today. The practice of making the assumptions in BC-IAMs testable could help to build confidence in their outputs, and filter out plausible from implausible structural assumptions.[6]

Third, policy choices should be based on estimates from many plausible, structurally distinct, models. As we have noted there are many important aspects of IAMs that we cannot hope to test empirically today, as the relevant verification data will only be realized long after current policies are enacted. There is thus substantial irreducible uncertainty about some of the core structural relationships in BC-IAMs. Exploring a wide range of structural assumptions – not just about overall technological change and economic growth, but also about climate damages and abatement costs – is crucial if we wish the policy prescriptions from modeling exercises to more accurately reflect the extent of our uncertainty about the consequences of climate policies. In our view, and that of others [5, 29], the set of BC-IAMs that are commonly applied in current policy analysis may underestimate the risks of inaction on climate change, in part because of a comparative lack of structural heterogeneity.

Fourth, the decision tools that are used to select policies should reflect the fact that our models are at best tentative predictive tools. Modern decision theory has developed a rich suite of tools for rational decision-making under deep uncertainty that allow us to express our lack of confidence in model output, yet most policy analysis with BC-IAMs still relies on decision tools that treat the uncertainty in climate policy as if it were of the same character as tossing a coin or rolling a die [30]. We should instead accept the limits of our knowledge, and use decision tools that fit the profoundly uncertain task at hand.

ACKNOWLEDGMENTS. AM acknowledges support from the Centre for Climate Change Economics and Policy, funded by the ESRC, and the Grantham Foundation for the Protection of the Environment. TM acknowledges support from the UCC Strategic Research Fund.

1. Oreskes N, Shrader-Frechette K, Belitz K, Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences (1994), Science, 263:641–646 .
2. Flato G, et al., Evaluation of Climate Models, Chapter 9 in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergonvermental Panel on Climate Change (2013), T. Stocker, et al., eds., Cambridge University Press, Cambridge, UK.
3. Knutti R, Should we believe model predictions of future climate change? (2008), Phil Trans Roy Soc A, 366:4647–4664.
4. Clarke L, et al., Assessing transformation pathways, Chapter 6 in Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (2014), O. Edenhofer, et al., eds., Cambridge University Press, Cambridge, UK.
5. Stern N, The Structure of Economic Modeling of the Potential Impacts of Climate Change: Grafting Gross Underestimation of Risk onto Already Narrow Science Models (2013), J Econ Lit, 51:838–859.

[5] Empirical tests of endogenous growth models suggest they are often more difficult to reconcile with historical data than simpler neoclassical alternatives [26, 27]. Section 5 of the Supporting Information describes our attempt to perform a similar model confirmation exercise on two endogenous growth models recently suggested as alternatives to DICE's model of TFP growth [28]. We find that both models are poorly behaved, being either not specified in a manner amenable to empirical estimation, or exhibiting instabilities that cause their predictions to be wildly uncontrolled when their parameters are estimated from historical data.

[6] [25] presents a further example of the benefits of hindcasting in a model of US energy intensity.

6. Pindyck R, Climate Change Policy: What Do the Models Tell Us? (2013), J Econ Lit 51:860–872.
7. Millner A, Dietz S, Heal G, Scientific Ambiguity and Climate Policy (2013), Env Res Econ 55:21–46.
8. Nordhaus W, Climate Clubs: Overcoming Free-Riding in International Climate Policy (2015), Amer Econ Rev, 105:1339–1370.
9. Grubb M, Köhler J, Anderson D, Induced technical change in energy and environmental modeling: Analytical approaches and policy implications (2002), Ann Rev Energy Enviro 27: 271–308.
10. Interagency working group on the Social Cost of Carbon, Technical support document: Technical update of the social cost of carbon for regulatory impact analysis - under executive order 12866, Tech. rep., United States Government (2013).
11. Roe G.H., Baker M. B., Why Is Climate Sensitivity So Unpredictable? (2007), Science 318:629–632.
12. Nordhaus WD, Estimates of the Social Cost of Carbon: Background and Results from the RICE-2011 Model (2011), Working Paper 17540, National Bureau of Economic Research.
13. Anderson B, Borgonovo E, Galeotti M, Roson R, Uncertainty in Climate Change Modeling: Can Global Sensitivity Analysis Be of Help? (2014), Risk Analysis 34:271–293.
14. Gollier C, Pricing the Planet's Future: The Economics of Discounting in an Uncertain World (2012), Princeton University Press.
15. Nordhaus WD, Sztorc P, DICE 2013: Introduction and User's Manual (2013).
16. Hope C, The marginal impact of CO2 from PAGE2002: An integrated assessment model incorporating the IPCC's five reasons for concern (2006),Integrated Assessment, 6:19–56.
17. Tol R, On the optimal control of carbon dioxide emissions: an application of FUND (1997), Environmental Modeling and Assessment, 2:151–163.
18. Golosov M, Hassler J, Krusell P, Tsyvinski A, Optimal Taxes on Fossil Fuel in General Equilibrium (2014), Econometrica 82:41–88.
19. D. Acemoglu, Introduction to Modern Economic Growth (Princeton University Press, 2008).
20. Gillingham K, Newell RG, Pizer WA, Modeling endogenous technological change for climate policy analysis (2008),Energy Econ 30, 2734–2753.
21. Piketty T, Zucman G, Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010 (2014), Quart J Econ 129:1255–1310.
22. Gordon RJ, Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds (2012), NBER Working Paper.
23. Iverson T, Denning S, Zahran S, When the long run matters (2015), Climatic Change, 129:57–72.
24. Prescott EC, Needed: A Theory of Total Factor Productivity (1998), Int Econ Rev 39:525–551.
25. Dowlatabadi H, Oravetz MA, US long-term energy intensity: Backcast and projection (2006), Energy Policy, 34(17):3245–3256.
26. Pack H, Endogenous Growth Theory: Intellectual Appeal and Empirical Shortcomings (1994), J Econ Persp 8:55–72.
27. Jones CI, Time Series Tests of Endogenous Growth Models (1995), Quart J Econ 110:495–525.
28. Dietz S, Stern N, Endogenous Growth, Convexity of Damage and Climate Risk: How Nordhaus' Framework Supports Deep Cuts in Carbon Emissions (2015), Econ. J, 125(583):574:620.
29. Revesz RL, et al., Global warming: Improve economic models of climate change (2014), Nature, 508:173–175.
30. Heal G, Millner A, Uncertainty and Decision Making in Climate Change Economics (2014), Rev Env Econ Policy, 8:120–137.
31. Nordhaus WD, Economic aspects of global warming in a post-Copenhagen environment (2010), Proc Nat Acad Sci, 107:11721–11726.
32. Easterly W, Levine R, What have we learned from a decade of empirical research on growth? It's Not Factor Accumulation: Stylized Facts and Growth Models (2001), World Bank Econ Rev 15:177–219.
33. Parente SL, Prescott EC, A Unified Theory of the Evolution of International Income Levels (2005), Handbook of Economic Growth, Steven N. Durlauf, Philippe Aghion, eds., vol. 1, Part B, 1371–1416.
34. Nordhaus WD, The DICE Model: Background and Structure of a Dynamic Integrated Climate-Economy Model of the Economics of Global Warming (1992), Cowles Foundation Discussion Paper 1009, Cowles Foundation for Research in Economics, Yale University.
35. Nordhaus WD, Yohe GW, Changing Climate: Report of the Carbon Dioxide Assessment Committee (1983), National Academies Press, 87–153.
36. Quandt RE, Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes (1960), J. Am. Stat. Assoc., 55:324–330.
37. Stock JH, Watson M, Introduction to Econometrics (2011), Pearson Education, Boston, Mass., Third Edition.
38. Brynjolfsson E, McAfee A The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies (2014), (W. W. Norton & Company, New York.
39. Acemoglu D, Aghion O, Bursztyn L, Hemous D, The Environment and Directed Technical Change (2012), Amer Econ Rev 102:131–166.
40. Romer P, Increasing Returns and Long-Run Growth (1986), J. Pol. Econ. 94:1002–1037.

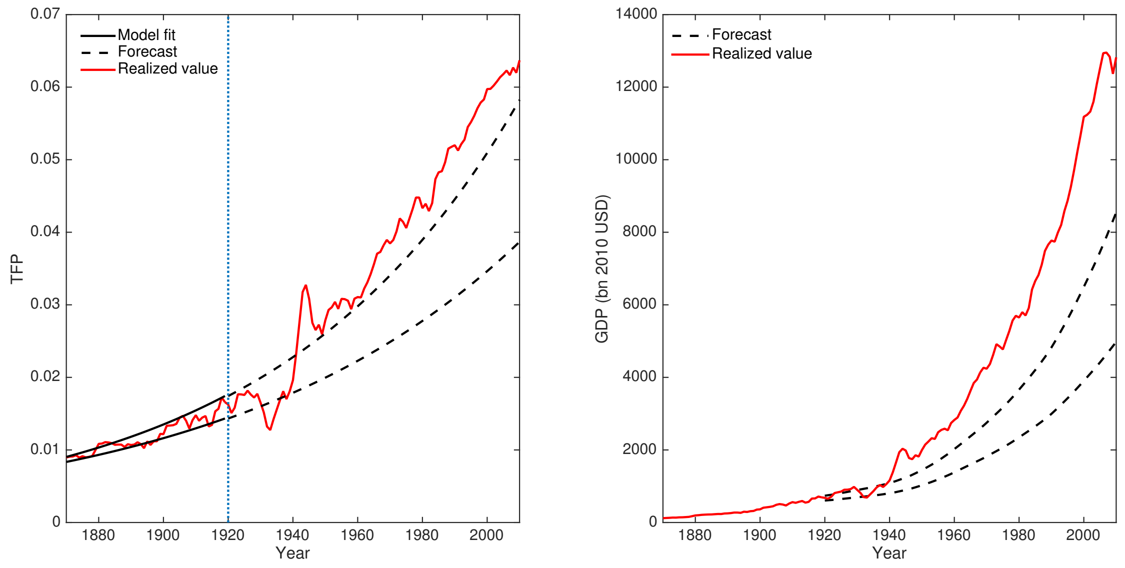

Fig. 1. Long-run forecasts of US total factor productivity (left panel) and economic output (right panel) from the economic growth model used by DICE. The model was trained on the data from 1870-1920, and projections made in 1920. Solid red lines are realized data values. Dashed black lines are forecasts generated using model parameter values at the boundary of estimated 95% confidence intervals, and solid black lines (left panel) show the fit of the TFP model to the data over the training window.

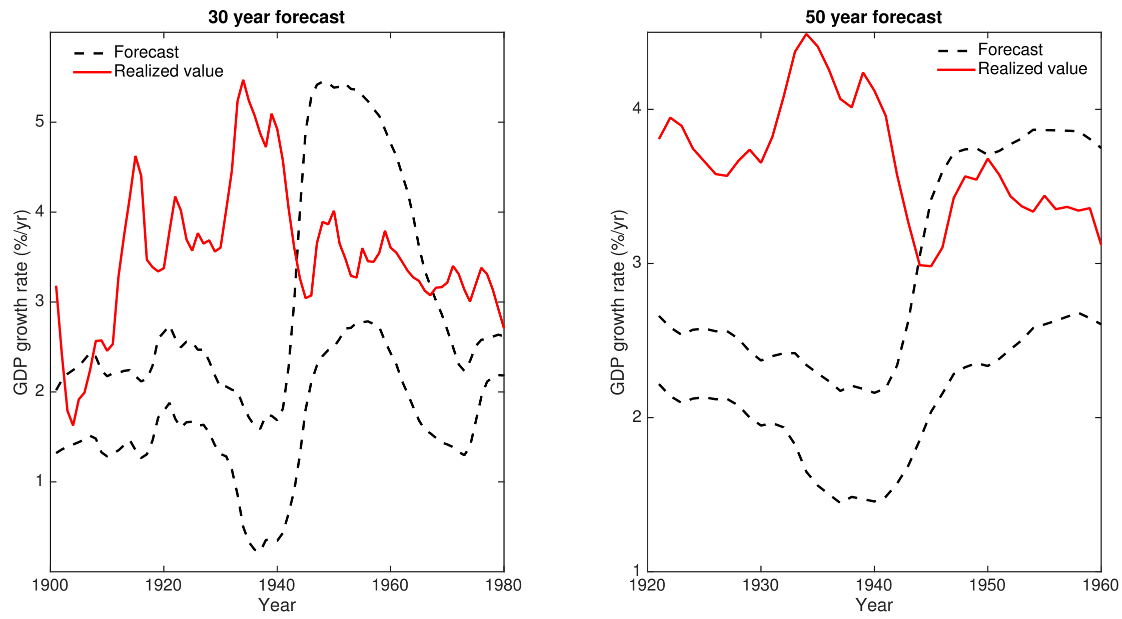

Fig. 2. 30 year (left panel) and 50 year (right panel) forecasts of US economic growth from the economic growth model used by DICE. For the left panel the solid red line at date T denotes the realized compound annual growth rate over the period $[T, T+30]$. The dashed black lines at date T denote the forecasted interval for the growth rate over $[T, T+30]$ when the model is trained on data from $[T-30, T]$, using the same estimation and prediction methodology as in Fig. 1. The right panel depicts the equivalent for 50 year training and verification windows.