National pathways to the Sustainable Development Goals (SDGs): A comparative review of scenario modelling tools

Article *in* Environmental Science & Policy · December 2016

DOI: 10.1016/j.envsci.2016.09.008

3 authors:

Cameron Allen
UNSW Sydney

Graciela Metternicht
UNSW Sydney

Thomas Wiedmann
UNSW Sydney

Cite as: Allen, C., Metternicht, G. and Wiedmann, T., 2016. National pathways to the Sustainable Development Goals (SDGs): A comparative review of scenario modelling tools. Environmental Science & Policy, 66, pp.199-207.

National pathways to the Sustainable Development Goals (SDGs): a comparative review of scenario modelling tools

Cameron Allen[1,*], Graciela Metternicht[2,†], Thomas Wiedmann[3,4†]

[1] Interdisciplinary Environmental Studies (IES), School of Biological, Earth and Environmental Sciences, UNSW Australia, Sydney, NSW 2052, Australia. E-mail: Cameron.allen@abs.gov.au

[2] IES, School of Biological, Earth and Environmental Sciences, UNSW Australia, Sydney, NSW 2052, Australia. E-mail: g.metternicht@unsw.edu.au

[3] Sustainability Assessment Program (SAP), School of Civil and Environmental Engineering, UNSW Australia, Sydney, NSW 2052, Australia. E-Mail: t.wiedmann@unsw.edu.au

[4] ISA, School of Physics A28, The University of Sydney, NSW 2006, Australia

***** Author to whom correspondence should be addressed; E-Mail: cameron.allen@abs.gov.au

Abstract:

The recently-adopted global Sustainable Development Goals (SDGs) will have significant implications for national development planning in both developed and developing countries in the post-2015 period to 2030. Integrated, nationally-owned SDG strategies will be at the centre of national efforts to implement the new sustainable development agenda. The long-run processes and systems perspective that are inherent in the SDGs present complex analytical problems for policymakers and analysts. Scenario analysis and quantitative modelling will be important analytical tools to support national sustainable development planning, and there is an increasingly sophisticated suite of models available to decision makers. This paper reviews and assesses a broad range of different quantitative models that have the potential to support national development planning for the SDGs. The study develops a typology and inventory of 80 different models, and then reviews the comparative strengths, weaknesses and general utility of different models through an initial screening and subsequent multi-criteria analysis of short-listed models. Current gaps in model capabilities are highlighted in the context of providing analytical support for national development planning for the SDGs. While some existing models are particularly relevant, it is unlikely that an ideal model can analyse all SDG targets and variables of interest within a single modelling framework. Top-down 'macro framework' models are likely to be more useful for undertaking system-level or economy-wide scenario analysis driven by the national long-term goals and targets, and for exploring trade-offs and synergies among sectors. Bottom-up sectoral models will be able to support far more detailed option-level impact analysis of concrete interventions, technologies and investments. Combining both approaches within an analytical framework will provide a robust approach for analysis and decision-making. The results highlight a range of potential gaps in current modelling capabilities, and provide some new tools to assist with model selection.

Keywords: scenario modelling; Sustainable Development Goals; sustainability science; ecological economics; multi-criteria analysis; Analytical Hierarchy Process

1. Introduction

1.1 The Sustainable Development Goals

In September 2015 the global community adopted the 2030 Agenda for Sustainable Development, which includes a set of 17 Sustainable Development Goals (SDGs) and 169 targets. The new goals aim to build on the achievements of the Millennium Development Goals (MDGs) and set a transformative agenda that emphasizes integration and balance among economic, social and environmental aspirations.

The SDGs are set to become a critical component of the new international development framework for all countries and will have major implications for national development planning efforts in the post-2015 period. While the goals themselves will be universal, it will be left to countries to select national targets and ultimately determine their own priorities and level of ambition in terms of the scale and pace of transformation. Integrated, nationally-owned SDG strategies will be at the centre of national efforts.

Compared with the MDGs, the SDGs represent a far broader and more integrated, complex and challenging agenda for countries to implement. They also apply to both developing and developed countries. The timeframe for achieving the goals will be medium- to long-term, with most goals and targets corresponding to a 2030 time horizon.

Governments will face many challenges in choosing realistic yet ambitious national targets and setting out the most cost-effective and appropriate pathways towards achieving them. Given the broad scope of the SDGs, policy-makers will need to be able to easily assess the economic, social and environmental implications of their strategies in an integrated way over the long-term.

With national implementation of the SDGs commencing in early 2016, this paper explores the strengths and weaknesses of contemporary modelling tools and gaps in current modelling capabilities in the context of national development planning for the SDGs, using a novel policy screening and criteria-based approach.

1.2 Modelling approaches and types of models: challenges in the context of the SDGs

In response to the perceived limitations of traditional approaches to economic development planning based on cost-benefit analysis and macroeconomic modelling (Barker, 2004; Daly and Farley, 2011; DeCanio, 2003; Scrieciu, 2007; Söderbaum, 2008), research activities in several disciplines have increasingly aimed at developing quantitative tools for the analysis of trade-offs and synergies among the three sustainable development dimensions. These approaches build upon a long tradition of modelling in economics, engineering and other disciplines which has resulted in the emergence of a suite of more sophisticated tools that are helping decision-makers to deal with the complexity that is inherent in sustainable development (Connolly et al., 2010; Howells et al., 2013; Joshi et al., 2015; Pedercini, 2011; Turner et al., 2011).

Scenario analysis and modelling has emerged as a method that is particularly well-suited to the task of taking a long-term view and attempting to harmonize socioeconomic and environmental goals (Kok et al., 2007; Miller et al., 2014; Raskin et al., 2010). Combined with quantitative modelling, it has become a widely-used approach for exploring possible or plausible future pathways and their potential outcomes and implications (Hertwich et al., 2015; Schandl et al., 2015; Swart et al., 2004; Vergragt and Quist, 2011).

Despite these advancements, policy-relevant modelling for sustainable development remains a challenge, and the gap between the outputs from commonly-used economic growth models and the advice needed to support decision making for the SDGs may be extremely large. To understand this gap, further research on the strengths and weaknesses of different modelling approaches and available models is needed.

Top-down modelling approaches can support national-scale, macro-framework analyses of interactions and feedbacks across a range of sectors; whereas bottom-up approaches can assist in detailed technological assessment and meaningful evaluation of alternatives at the sectoral level. Combining these into hybrid approaches facilitates complementary analyses that balance the strengths and weaknesses of different approaches (Hourcade et al., 2006; van Vuuren et al., 2009)

The 'types' of models is another aspect that needs consideration in national policy for sustainable development. Input-output models are a useful descriptive tool for a national economy and form the basis for many advanced models, however their static nature limits their value in terms of long-term scenario modelling (Catenazzi, 2009; Herbst et al., 2012). Macro-econometric models are dynamic and based on a large amount of historical data, though they are of limited value for long-term analysis (Hedenus et al., 2013; Pollitt et al., 2010; Van Beeck, 1999). Dynamic computable general equilibrium (CGE) models are well-suited for scenario analysis using a consistent theoretical framework and feedbacks across sectors, however their theoretical underpinnings may render them less suitable for modelling sustainable development transitions (Barker, 2004; Bhattacharyya and Timilsina, 2010; Scrieciu, 2007). System dynamics models are also suited to all types of scenario analysis and comparative assessment of alternatives, however the definition of correct boundaries and feedback loops can be problematic (Hjorth and Bagheri, 2006; Nicholson, 2007; Pedercini, 2003, 2011; Turner et al., 2011). Bottom-up optimisation and simulation models are more useful for sector-based planning due to their more limited scope and detailed coverage of technologies and alternatives, however they generally lack feedback loops with other sectors in the broader economy (Herbst et al., 2012; Nicholson, 2007; Pollitt et al., 2010). Multi-agent models are promising in the context of sustainable development (Boulanger and Bréchet, 2005), though they are highly complex and remain experimental, with limited practical application (Wieland and Gutzler, 2014). Finally, hybrid and integrated assessment models are leveraging the strengths of multiple modelling methodologies, addressing some of the weaknesses related to the aforementioned categories and providing a more flexible and tailor-made approach (Bazilian et al., 2011; Herbst et al., 2012; Hourcade et al., 2006).

This evidences the absence of unique modelling approaches and model types that will address all analytical requirements required by underpinned by the new SDGs. Rightly, Nicholson (2007) points out the lack of a single universal methodology suited to all problems; with the utility of modelling approaches dependent on the nature of the system of focus, and the type of prediction desired. Likewise, the choice of modelling tools depends on the priority sectors of concern, the availability of expertise and data, cost and time limitations, amongst other factors (Böhringer and Löschel, 2006; Börjeson et al., 2006; Höjer et al., 2008).

1.3 Model types: state-of-the-art

A considerable volume of academic literature has emerged over the past ten years as international interest in models has grown along with their level of sophistication and computing power. This has included a number of evaluative or comparative reviews of different types of models in the context of a specific policy issue or research question, including sustainable development. For a brief summary of 40 past reviews of models drawn from the literature, refer to **Table A** in the Supplementary Material (SM).

While not exhaustive, this provides a useful starting point for an initial comparative analysis of models in the context of the SDGs. A review of this literature highlights several research gaps. Firstly, with the emergence of a large volume of modelling tools adopting various methodologies over the past decade, there is a lack of a consolidated inventory or list of models from which decision makers can select a tool that best suits their analytical needs, and which facilitates an informed choice.

Secondly, there is no standardized approach for categorizing the various models, which makes it difficult to explore model strengths and weaknesses and provide an intuitive framework for such an inventory. As Nicholson (2007) again points out, “there exist a plethora of model types, and perhaps only a slightly smaller number of model classification schemes”.

Thirdly, while the literature provides numerous critiques of specific models as well as broad model categories, there is limited consideration of the utility of different models to support national sustainable development planning. Finally, there is no literature at this stage that considers these questions specifically in the context of the new global SDGs.

The methodologies used in the literature to comparatively assess different models generally fall into two approaches: comparative assessment based on key desirable attributes, characteristics or criteria; or, less often, comparative assessment of model outputs and projections. As pointed out by Wieland (2012), systematic comparisons of the empirical outputs of a large variety of available models are rare because multiple model runs and simulations are typically resource-intensive and costly. This approach is more commonly used where a small number of models are being reviewed in the context of a very specific policy question.

2. Methodology

The objective of this study was to undertake an initial comparative assessment of a broad range of models and their relevance for national SDG planning. As such, it was not considered feasible to run detailed model simulations and compare results; instead, a multi criteria assessment (MCA) approach was adopted as outlined below.

2.1. Development of a typology for distinguishing among and categorizing different models (Step 1)

Initially, a systematic review of over 40 academic journal articles and research papers from the modelling literature was undertaken with the aim of developing a ‘typology’ of models based on the most common ‘distinctions’ used by experts to differentiate among models (**Table A**, SM). For each article, the main distinctions used were recorded in a template, which was then analysed to identify the most common distinctions and categories used by authors. This analysis provided a consolidated, consistent and robust model typology covering a total of ten model distinctions (see **Table B**, SM). In particular, the previous work of Van Beeck (1999) is acknowledged.

2.2. Inventory of modelling options currently used in national development planning (Step 2)

The typology served as the framework for a tabular inventory of modelling options which was then populated with models identified in the literature and other sources (e.g. best practice toolkits, national assessment reports, strategy documents, and the web). The review identified a broad set of 80 models that were used by analysts to support national development and sustainable development planning as well as some well-known global scenario models that were commonly referenced in the literature (see **Table H,** SM). The aim was to include a sufficiently broad range of models across different model categories and, in particular, those most commonly and recently used in different national contexts.

2.3. Formulation, validation and weighting of assessment criteria and expert survey (Step 3)

To undertake the comparative analysis, twelve criteria were initially developed based on a review of criteria used in previous comparative assessments of models (**Table A**, SM) as well as key attributes of the SDGs of relevance for modelling (**Table C**, SM). The criteria were incorporated into a survey and subjected to an expert validation and weighting process. The survey was circulated to 70 experts from the academic and research community as well as national development planning and modelling professionals, who were asked to score the importance of each criteria, as well as to validate each criteria in terms of completeness, redundancy, 'operationality', double counting, and size (Dodgson et al., 2009). This step resulted in a final set of ten revised and weighted criteria for use in the comparative assessment.

2.4. Initial screening of the broad inventory of models and identification of SDG thematic issues addressed (Step 4)

Given the large number (80) and diversity of models to be assessed, an initial screening of the models was undertaken to remove models considered redundant or less relevant for SDG planning. The screening exercise was undertaken based on publicly available model documentation, websites, manuals and academic literature. It is acknowledged that the available information for some models was limited. The screening was based on the two criteria that received the highest scores through the expert survey: i.e. 'integrated' and 'policy relevant'. Each model in the inventory was also reviewed to identify which thematic policy issues (i.e. which of the 17 SDGs) they were capable of analysing (**Table G**, SM). Based on this initial screening, eight models were short-listed for a detailed comparative assessment.

2.5. Final comparative assessment of short-listed modelling options and calculation of rankings (Step 5)

MCA of the eight short-listed models was undertaken based on the full set of ten criteria. The assessment used the Analytical Hierarchy Process (AHP) which is a MCA scoring methodology based on pairwise comparison. AHP is used to convert a subjective assessment of relative importance into a set of overall scores or weights, and it suits circumstances where judgements, rather than measurements, are the predominant form of input information (Saaty, 1990). The methodology in this study was informed by previous comparative assessments in the literature, including Boulanger and Bréchet (2005).

Using a standard AHP preference scale (**Table D**, SM), the comparative effectiveness of each of the different models in meeting the ten criteria was estimated.

Formally, we let $M = [M_1, M_2, \ldots, M_m]$ be the set of m models to be assessed, and $C = [C_1, C_2 \ldots, C_n]$ the set of n criteria. For each criterion (C_z), each model (M_x) was compared to each other model (M_y) using the preference scale.

For example, if the model M_x was considered slightly better than model M_y with respect to the criterion C_z, then its score was (α_{xyz}) = 3. Because this is a reciprocal relationship, model M_y is automatically rated lower than model M_x, with a score (α_{yxz}) = 1/3. The output of this analysis results in a matrix for each criterion. The last column in each matrix shows the geometrical mean (GEOMEAN) of the row values which is generally a good approximation of the principal eigenvector as per the standard AHP methodology. This eigenvector expresses the 'relative adequacy' of the model to each criterion. The more a model outperforms other models, the higher the value.

The relative adequacy scores from the AHP analysis were combined with the 'importance' weightings that were given to each criterion through the expert survey. The relative weights for each criterion (W_{cx}) were calculated using the following basic formula:

$$W_{cx} = S_{cx} / \sum(S_{c1}, S_{c2}, S_{c3} \ldots S_{c10}) \quad (1)$$

For example, the relative weight of criterion W_{c1} was equal to its score S_{c1} divided by the sum of scores for all criteria ($S_{c1}, S_{c2}, S_{c3} \ldots S_{c10}$). Relative weights for each criterion were then multiplied by scores for each model to give final weighted scores and rankings.

The evidence base used to make this assessment included: 1. model technical manual, user guide and/or developer documentation; 2. model website; 3. academic and expert literature citing the model; 4. assessments, reports and strategies/plans that provide case studies of model application; and 5. test runs of the model software on a personal computer (where available).

3. Results and Discussion

3.1. Development of model typology

Ten distinctions for differentiating among models were identified in the model typology (**Table B**, SM), which provided the framework for the inventory of models. The two distinctions most commonly used in the literature were the ‘analytical approach’ and the ‘model category/type’. Three analytical approaches were identified, based on whether the model adopted a ‘top-down’, ‘bottom-up’ or ‘hybrid’ approach. Building on this, eight modelling categories were identified: (1) top-down input-output models; (2) top-down econometric models; (3) top-down CGE models; (4) top-down system dynamics models; (5) bottom-up optimization/partial equilibrium models; (6) bottom-up simulation models; (7) bottom-up multi-agent models; and (8) hybrid models (various combinations of the previous categories).

3.2. Inventory of modelling options

The largest number of models in the inventory (**Table H**, SM) fell into the hybrid model ‘category’ (39%), followed by the top-down CGE models (19%) (**Figure 1**). There are several likely reasons for this, including that hybrid models are more suited to complex sustainable development problems, that there has been an increased uptake of such models in recent national sustainable development planning, that they have received more attention in the academic literature due to new innovations, and that the hybrid category itself encompasses a very large variety of combinations of models.

The models reviewed were primarily dynamic in nature (91%) and adopted both a simulation (60%) and optimisation (55%) model solution. Many of the hybrid models integrated both of these modelling solutions into a single framework. A majority of the models also targeted the national scale (71%) and supported long-term analysis (80%). These will be important attributes needed for models to support national development planning for the SDGs, as discussed in more detail hereafter.

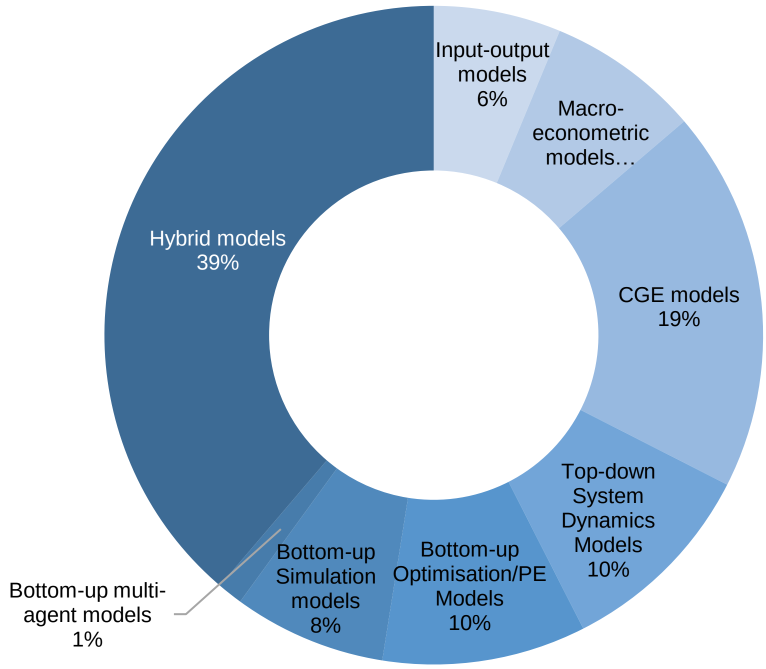

Figure 1. Percentage of the 80 models reviewed falling into each of the eight 'model categories' identified.

3.3. Screening and MCA of models for national development planning for the SDGs

The subsequent screening of the 80 models and MCA of eight short-listed models provided the means for a more in-depth, comparative assessment of the models in the context of national development planning for the SDGs.

3.3.1 Formulation, validation and weighting of criteria

The final set of ten criteria are listed in **Table 1**, including the average scores allocated by experts through the survey. Detailed descriptive text was prepared for each of these criteria to assist with interpretation in the expert survey and for application in the comparative assessment (see **Table E**, SM). Feedback from experts during the survey helped refine the final set of criteria and descriptive text, and also influenced the approach used in the subsequent comparative assessment:

- There is unlikely to be a single model that will be able to meet all of the ten criteria or address all of the priority thematic policy issues covered by the SDGs. As the coverage of issues addressed by a model is broadened, the analysis inevitably loses depth, and ultimately decision makers will require both. A model should be focused on priority, interrelated issues rather than trying to address every issue.
- The selection of a particular modelling tool needs to reflect the priority SDGs and targets of interest for a particular country, and this needs to be factored into the comparative analysis.
- The relatively large number of criteria will make it difficult to apply to a large set of 80 different models, particularly given the great diversity among the different modelling tools.

Table 1. Final set of criteria and associated weightings.

Criteria Related to Scientific Strength	Score	Weight
1. Integrated/Inter-linkages/Inter-disciplinary	8.73	0.118
2. Dynamic & long-term perspective	8.06	0.109
3. Systems-based yet realistic/meaningful	7.75	0.104
4. Transformative	7.00	0.094
5. Global-local perspective	6.63	0.089

6.	Participatory, transparent and legitimate	6.50	0.088
Criteria Related to Model Application and Usage		**Score**	
7.	Policy relevance, scenario analysis, & policymaking guidance	8.60	0.116
8.	Applications, visibility, flexibility – in both developed & developing countries	7.40	0.100
9.	Ease of use/user friendliness	7.20	0.097
10.	Cost, time and effort (for model development	6.40	0.086

NB: Scores (out of a maximum of 10) represent averages of expert ratings. Weightings were calculated using Formula (1).

3.3.2. Initial screening of modelling options

The table produced through the model screening exercise (**Table G,** SM) indicates whether or not the 80 models reviewed met each of the two screening criteria as well as their relevance in terms of modelling each of the 17 SDGs. This screening table is a useful tool in itself that can assist decision makers in identifying models that are more likely to be suitable for national SDG planning and that address priority sectors of national interest (e.g. relating to energy, climate change, water, poverty, jobs, biodiversity, or any combination thereof).

It is important to recognize that all of the models reviewed serve a purpose and could play a role in policy analysis for the SDGs at different scales and with different policy questions in mind. As such, they should not be disregarded altogether by decision makers. As pointed out by the experts, there will not necessarily be a single ideal model that is capable of modelling all of the SDGs.

From the total of 80 models reviewed, only eight models (10%) met the two screening criteria of 'policy relevant' and 'integrated'. The shortlist comprised one top-down CGE model (*MAGNET*), one top-down system dynamics model (*Polestar*), and six hybrid models (*IMPACT*; *International Futures*; *Threshold 21*; *EC4MACS*; *InVEST*; and *LowGrow*).[1] This highlights the greater potential utility of hybrid models in supporting analysis for the SDGs.

Interestingly, inability to meet the 'integrated' criteria resulted in most models being screened out. The integrated criteria related to whether or not a model incorporated variables relating to the economic, social and environmental dimensions in their modelling framework. From the total of 80 models reviewed, only 15 models (19%) were considered to have generally met this criterion, meaning that most models (over 80%) failed to incorporate variables from all three dimensions in their modelling framework. This highlights a gap in modelling capabilities for the SDGs.

However, the limitations of this screening criterion are also acknowledged in that it tended to screen out technology-rich bottom-up models that focus on a specific sector and lack system-level feedbacks to broader socioeconomic and environmental variables. In practice, such models could be coupled with top-down analysis to ensure that these feedbacks, trade-offs and synergies are adequately addressed.

In the context of planning for the SDGs, bottom-up sectoral models with high technological detail could support far more detailed option-level impact analysis of concrete interventions that can better assist policy formulation at the sectoral level. This could be complemented by macro-framework top-down models that address a broad range of SDGs, undertaking a system-level impact analysis driven by the long-term goals and targets, and exploring different scenarios or pathways to assess the scale and pace of change required across different sectors, and interlinkages and feedbacks among sectors.

In terms of sustainability dimensions, the most common approach was the integration of economic and environmental variables in the model **(Figure 2)**. A total of 54 models (or 68%) integrated these two dimensions to varying degrees, highlighting the rapidly growing catalogue of economy-environment models. The social

[1] For a short description of these models, see model inventory in **Table H** (SM).

dimension of sustainable development is by far the least addressed, with only 17 models (or 21%) including social variables within their modelling framework, and often with very limited coverage (most commonly a few health-related or nutrition-related variables). The modelling of social variables can therefore also be considered as an important gap in modelling capabilities.

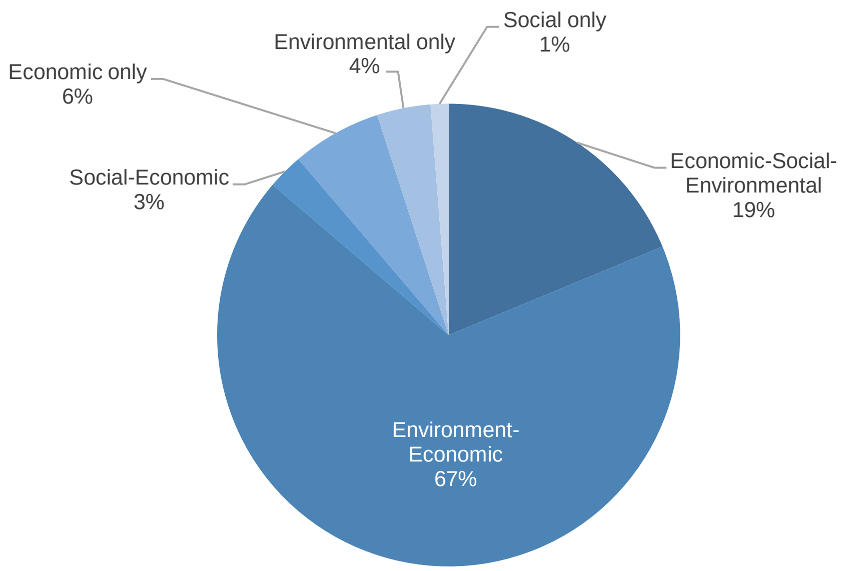

Figure 2. Sustainable development dimensions addressed by the 80 models reviewed. The figure shows the percentage of the models reviewed that addressed the social, economic and environmental dimensions (individually or in combination).

The models reviewed tended to perform better in terms of the 'policy relevant' screening criterion. In order to meet this criterion, a model had to be dynamic, able to support a medium-to-long-term planning horizon of at least 15 years, and applicable for a national scale analysis with evidence that it had been used for this purpose in previous studies. Of the total of 80 models, 47 (or 59%) were considered to meet this criterion. For those that did not, the most common reason was that the spatial resolution of the model meant that it was not sufficiently capable of national-scale analysis. For example, 19 of the models reviewed focused on the global level and it was considered they had inadequate downscaling to support meaningful national-scale analysis.

In terms of geographic focus, many of the models tended to focus on a specific country or region (most commonly the European Union) where the models were developed and have been extensively applied. Several models also focused specifically at the national scale and had only been used in a single country (e.g. Australia, Canada and the USA). A key consideration for these models is whether or not they are transferrable to other country contexts. The remaining models would appear to be more flexible and adaptable, with no specific country or regional emphasis.

The vast majority (89%) of the models reviewed can be considered dynamic models that are capable of analysis over at least a 15-year timeframe. A total of five models were considered as static, mainly input-output databases or older economic models. Another four models were suited to shorter-term analysis only, mainly macro-econometric models.

The review also evidences that model coverage varies substantially across the SDG policy areas. The models with the greatest coverage of policy issues were also generally included in the group of eight shortlisted models, in particular *International Futures* and *Threshold 21*. Of all the models reviewed, only *International Futures* included variables relating to all of the thematic issues addressed by the 17 SDGs, however variables were very limited for many goals.

The SDGs with the greatest coverage in terms of modelling capacity were economy and employment (SDG 8), energy (SDG7), investment-trade-finance (SDG 17), food-land (SDG 2), climate change (SDG 13), and to a lesser degree, water (SDG 6) and industry-innovation (SDG 9) (**Figure 3**). Policy areas that represent key gaps in modelling capabilities included institutions-governance (SDG 16), gender equality (SDG 5), cities (SDG 11), marine life (SDG 14), education (SDG 4) and, to a lesser degree, poverty (SDG 1), health (SDG 2), inequality (SDG 10) and terrestrial life (SDG 15).

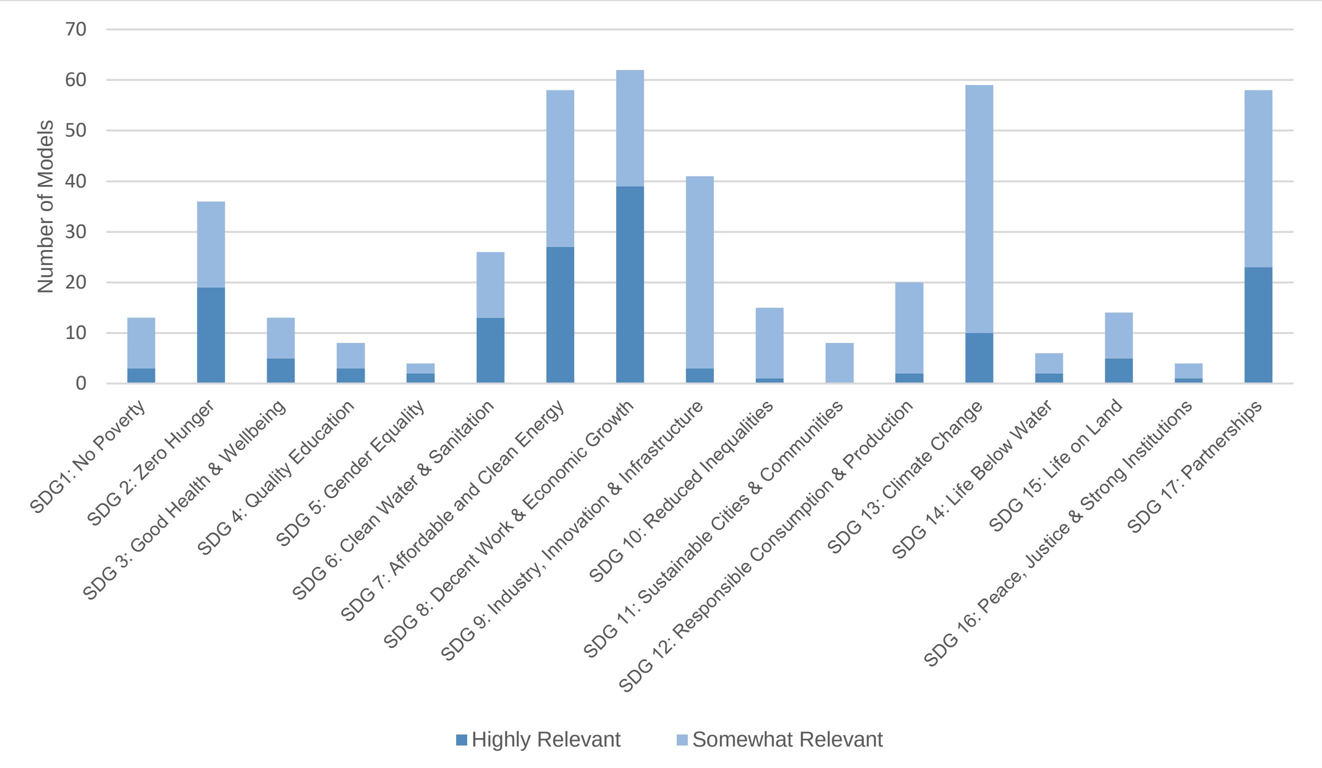

Figure 3. Model coverage of SDG thematic policy areas. The figure shows the number of models that were reviewed that were considered ‘highly relevant’ or ‘somewhat relevant’ for analysing a particular thematic policy area.

Table 2 highlights the particularly poor coverage of many social issues and some environmental issues. However, for some categories (e.g. cities or sustainable consumption and production), this is also due to the lack of clarity regarding the variables covered by the goal, which would likely overlap with other goals (e.g. energy, water, climate change). Despite this ambiguity, key gaps or thematic issues requiring further model development include poverty, health, education, gender, inequality, sustainable consumption and production, biodiversity and governance-institutions.

Table 2. Coverage of SDG policy issues addressed by the inventory of models.

	SDG1: No Poverty	SDG 2: Zero Hunger	SDG 3: Good Health & Wellbeing	SDG 4: Quality Education	SDG 5: Gender Equality	SDG 6: Clean Water & Sanitation	SDG 7: Affordable and Clean Energy	SDG 8: Decent Work & Economic Growth	SDG 9: Industry, Innovation & Infrastructure	SDG 10: Reduced Inequalities	SDG 11: Sustainable Cities & Communities	SDG 12: Responsible Consumption & Production	SDG 13: Climate Change	SDG 14: Life Below Water	SDG 15: Life on Land	SDG 16: Peace, Justice & Strong Institutions	SDG 17: Partnerships
Highly Relevant	**3**	**19**	**5**	**3**	**2**	**13**	**27**	**39**	**3**	**1**	**0**	**2**	**10**	**2**	**5**	**1**	**23**
Somewhat Relevant	**10**	**17**	**8**	**5**	**2**	**13**	**31**	**23**	**38**	**14**	**8**	**18**	**49**	**4**	**9**	**3**	**35**

NB: Figures represent the number of models. The colour scale highlights those SDG policy issues with good (green), moderate (white), or poor (red) coverage). Lighter shades of these colours represent intermediate values.

3.3.3. Final comparative assessment: AHP, weighted scores and ranking of short-listed models

The AHP methodology used for the MCA of the eight short-listed models produced output matrices for each of the ten criteria (**Table F**, SM). Combining together the geometrical means (GEOMEAN) for each criterion provided a summary matrix representing comparative scores or preference ratings for each model across all criteria (**Table 3**). Overall rankings for each of the models were then allocated based on the GEOMEAN sums.

It can be seen that the two hybrid macro-framework models *Threshold 21* and *International Futures* performed the best overall in comparison with the other models across all ten criteria. Both of these models were consistently better than other models against almost all of the criteria, partly owing to their integrated and systems-based framework, broad policy scope, long-term perspective, high flexibility, transparency, user-friendliness and hybrid approach.

Threshold 21 scored the highest overall (30.6), partly due to its greater application/visibility and flexibility with multiple national applications in developing and developed countries (in particular for national green economy or MDG strategies), as well as the participatory process by which the model has been applied and its transparency in terms of model structure, methodology and assumptions. However, while an evaluation version of the model was available to review the software and interface, the model cannot be downloaded or purchased 'off-the-shelf'. Rather, it generally comes as part of a technical assistance project. The considerable size, scope and complexity of the model mean that resources, costs and time per application could be considered to be moderately high.

Table 3. Overall scores and ranking of shortlisted models using AHP.

	1. Integrated	2. Dynamic & Long-term	3. Systems-based & meaningful	4. Transformative	5. Global-local perspective	6. Participatory, transparent, legitimate	7. Policy relevance/ scenarios	8. Application, visibility, flexibility	9. User friendly	10. Cost, time and effort	GEOMEAN Sums
Threshold 21	2.920	2.546	3.807	3.840	1.914	4.456	2.565	4.671	3.060	0.799	30.6
LowGrow	0.550	0.252	0.562	0.594	0.256	0.202	0.387	0.151	1.052	0.799	4.8
MAGNET	0.293	1.130	0.278	0.210	0.357	0.256	0.387	0.848	0.205	0.389	4.4
Polestar	2.920	2.546	0.837	1.174	0.688	1.439	0.567	0.477	2.111	0.799	13.6
IMPACT	0.550	1.130	0.837	1.174	1.037	0.483	2.565	1.565	0.403	2.310	12.1
International Futures	2.920	2.546	2.918	3.840	2.697	2.030	1.233	1.565	3.060	3.286	26.1
InVEST	0.293	0.252	0.389	0.336	0.688	3.379	0.567	2.879	1.052	2.918	12.8
EC4MACS	1.550	0.748	2.115	1.174	4.318	0.909	2.565	0.497	0.554	0.227	14.7

NB: The colour scale indicates where the value for each criterion falls within the range of values across the eight models – dark green reflecting the highest *relative* score, dark red the lowest *relative* score, and yellow a middle-range *relative* score. Lighter shades of these colours represent intermediate values.

International Futures was ranked second overall (26.1), with comparable performance to *Threshold 21* across many of the scientific rigor criteria. However, the model was initially developed by researchers as an education tool and while it has been applied to many global studies, there was more limited evidence of application in national development planning. As such, it scored lower in terms of application/visibility, policy relevance, and participation/transparency. An advantage of the model over *Threshold 21* was that it is available and free to download, making it highly accessible and affordable, with training courses available for users at reasonable costs. However, as with *Threshold 21*, its considerable size and scope mean that time and resources required per application would also be high.

The lowest-scoring models were the macroeconomic CGE model *MAGNET* and the hybrid model *LowGrow*; they tended to score comparatively lower across all of the criteria due to several perceived shortcomings when compared against the other models. With regard to scientific rigor, the models exhibited a narrower focus, with limited integration and interlinkages, absence of a transformative or system-based approach, and adoption of a top-down macroeconomic modelling methodology with high aggregation of outputs. The *LowGrow* model has only been applied for research purposes in a single country (Canada) limiting its applicability, flexibility and policy relevance. The highly technical nature and complexity of the *MAGNET* model meant that it scored lower in terms of transparency, ease of use, and cost, time and effort for model development.

The collective performance of the models against each criterion was also reviewed in order to highlight potential strengths and weaknesses in model capabilities overall. In general, the short-listed models collectively performed better in terms of their dynamic model solutions, long-term modelling horizons, integration of variables across all three dimensions of sustainable development with interlinkages and feedback loops (though often for a limited range of goals and variables), and ability to undertake different types of scenario analysis (baseline, what-if, backcasting) and evaluate alternatives to inform policy. Key weaknesses or gaps were evident in terms of the cost, time and effort required for each model application and capacity to model different spatial scales, disaggregate data and provide spatially-explicit outputs.

Potential gaps and areas for improvement included: the extent to which models can evaluate transformations in terms of decoupling of material and resource consumption from economic development as well as impacts on human wellbeing; model transparency in terms of its development and application as well as its structure and assumptions; flexibility and adaptability of models and their diffusion in both developed and developing country contexts; and the user friendliness of models in terms of their interface, manuals, training and outputs. Balancing the need for a systems-based approach with the need for meaningful model outputs also remains a challenge.

The calculation of weighted scores had only a marginal impact on the overall rankings, with *Threshold 21* and *International Futures* retaining their first and second rankings. These integrated, systems-based macro-framework hybrid models clearly out-performed the other short-listed models across all criteria. As such, they are considered particularly applicable for scenario modelling to support national SDG planning. Final model rankings based on weighted scores were: 1. *Threshold 21*; 2. *International Futures*; 3. *EC4MACS*; 4. *Polestar*; 5. *IMPACT*; 6. *InVEST*; 7. *LowGrow*; and 8. *MAGNET.*

3.4 Limitations of the comparative assessment approach

It is important to note that this analysis has limitations. While the results help to illustrate that certain models may be particularly applicable for national SDG planning, it does not mean that they will be suited to every situation, or that they should be used in isolation. Nor does it mean that the other models reviewed will not contribute to national SDG planning. Different models will be more or less applicable depending upon the specific national context and decision makers will select a tool based on their specific circumstances and needs.

The comparative analysis using the ten criteria tended to favour those models that take a broad and integrated, systems-based approach, encompassing many SDGs and targets and their interlinkages and feedback loops. In addition, high-scoring models tended to have greater flexibility and adaptability, evidence of scenario modelling applications in different national settings, and well-developed and user-friendly interfaces with access to training and technical support. The analysis ranked as less suitable those models that tended to focus on a narrow set of SDG policy issues with limited linkages and feedbacks to broader systems, limited national application, and highly complicated, inaccessible or difficult-to-use software interfaces. The subjectivity of such an approach is acknowledged.

Based on these limitations, the final rankings may be best interpreted in the context of the model's relevance for supporting macro-framework scenario modelling for the SDGs, rather than bottom-up sectoral or technological modelling. In practice, such macro-analysis could and should be complemented by more detailed bottom-up modelling and analysis of specific alternatives at the sectoral level.

Selection of the most appropriate model (or complementary models) for a particular set of SDG policy priorities can be guided by the screening table developed as part of the study, which highlights the relevant sectors (i.e. SDGs) that can be analysed by each model (**Table G**, SM).

4. Conclusions

This review paper demonstrates that modelling capability for supporting implementation of the SDGs at national level exists, though selection of appropriate modelling approaches and model types will be dependent upon the priority sectoral issues of interest, amongst other factors. Accordingly, guidance has been presented in the form of a model inventory, screening table, and comparative analysis to assist with model selection.

As only one of the 80 models reviewed included variables relating to all 17 SDGs, it is concluded that there may not be a single model that can analyse all variables of interest, and that the most robust approach will likely be to limit the analysis to a nexus of priority sectoral issues, and combine different models that address these issues in a broader analytical framework

Overall, the strengths of the models reviewed included their dynamic nature, national scale and capacity to support long-term scenario analysis. These are important attributes needed for models to support national SDG planning. However, the lack of integration of variables across all dimensions of sustainable development is a perceived weakness, in particular the inability to model social variables.

Eight models are perceived to be of greater relevance for national SDG planning, collectively performing better in terms of their dynamic model solutions, long-term modelling horizons, integration of variables, and ability to undertake different types of scenario analysis and evaluate alternatives to inform policy. Both *Threshold 21* and *International Futures* clearly outperformed other shortlisted models in the comparative analysis, however they will not be suited to support modelling related to all SDGs.

Based on the identified gaps in modelling capabilities, priorities for further model development to support SDG implementation would include: ability of modelling at different spatial scales, data disaggregation, and providing spatially-explicit outputs (e.g. maps); the modelling of 'decoupling' relationships between resource consumption and economic development and human wellbeing; improving the user-friendliness of models and access to model software, training and technical support; increasing transparency in terms of model development and application as well as structure, assumptions and limitations; and enhancing flexibility and adaptability of models to the diverse contexts of developed and developing countries tasked with advancing the 2030 agenda for sustainable development.

Acknowledgments

We wish to acknowledge the contribution of the experts who were surveyed as part of the validation and weighting processes outlined in this study, as well as the anonymous reviewers for their valuable comments.

References

Barker, T., 2004. The transition to sustainability: a comparison of General–Equilibrium and Space–Time–Economics approaches. Tyndall Centre Working Paper.
Bazilian, M., Rogner, H., Howells, M., Hermann, S., Arent, D., Gielen, D., Steduto, P., Mueller, A., Komor, P., Tol, R.S., 2011. Considering the energy, water and food nexus: Towards an integrated modelling approach. Energy Policy 39, 7896-7906.
Bhattacharyya, S.C., Timilsina, G.R., 2010. A review of energy system models. International Journal of Energy Sector Management 4, 494-518.
Böhringer, C., Löschel, A., 2006. Computable general equilibrium models for sustainability impact assessment: Status quo and prospects. Ecological economics 60, 49-64.
Börjeson, L., Höjer, M., Dreborg, K.-H., Ekvall, T., Finnveden, G., 2006. Scenario types and techniques: towards a user's guide. Futures 38, 723-739.
Boulanger, P.-M., Bréchet, T., 2005. Models for policy-making in sustainable development: The state of the art and perspectives for research. Ecological Economics 55, 337-350.
Catenazzi, G., 2009. Advances in techno-economic energy modeling. ETH Zurich.
Connolly, D., Lund, H., Mathiesen, B.V., Leahy, M., 2010. A review of computer tools for analysing the integration of renewable energy into various energy systems. Applied Energy 87, 1059-1082.
Daly, H.E., Farley, J., 2011. Ecological economics: principles and applications. Island press.
DeCanio, S.J., 2003. Economic models of climate change: a critique. Palgrave Macmillan New York.
Dodgson, J., Spackman, M., Pearman, A., Phillips, L., 2009. Multi-criteria analysis: a manual. Department for Communities and Local Government: London.
Hedenus, F., Johansson, D., Lindgren, K., 2013. A Critical Assessment of Energy-economy-climate Models for Policy Analysis. Journal of Applied Economics and Business Research 3, 118-132.
Herbst, A., Toro, F., Reitze, F., Jochem, E., 2012. Introduction to energy systems modelling. Swiss journal of economics and statistics 148, 111-135.
Hertwich, E.G., Gibon, T., Bouman, E.A., Arvesen, A., Suh, S., Heath, G.A., Bergesen, J.D., Ramirez, A., Vega, M.I., Shi, L., 2015. Integrated life-cycle assessment of electricity-supply scenarios confirms global environmental benefit of low-carbon technologies. Proceedings of the National Academy of Sciences 112, 6277-6282.

Hjorth, P., Bagheri, A., 2006. Navigating towards sustainable development: A system dynamics approach. Futures 38, 74-92.
Höjer, M., Ahlroth, S., Dreborg, K.-H., Ekvall, T., Finnveden, G., Hjelm, O., Hochschorner, E., Nilsson, M., Palm, V., 2008. Scenarios in selected tools for environmental systems analysis. Journal of Cleaner Production 16, 1958-1970.
Hourcade, J.-C., Jaccard, M., Bataille, C., Ghersi, F., 2006. Hybrid Modeling: New Answers to Old Challenges, Introduction to the Special Issue of The Energy Journal. The Energy Journal, 1-11.
Howells, M., Hermann, S., Welsch, M., Bazilian, M., Segerström, R., Alfstad, T., Gielen, D., Rogner, H., Fischer, G., van Velthuizen, H., 2013. Integrated analysis of climate change, land-use, energy and water strategies. Nature Climate Change 3, 621-626.
Joshi, D.K., Hughes, B.B., Sisk, T.D., 2015. Improving Governance for the Post-2015 Sustainable Development Goals: Scenario Forecasting the Next 50 years. World Development 70, 286-302.
Kok, K., Biggs, R., Zurek, M., 2007. Methods for developing multiscale participatory scenarios: insights from southern Africa and Europe. Ecology and Society 13, 8.
Miller, T.R., Wiek, A., Sarewitz, D., Robinson, J., Olsson, L., Kriebel, D., Loorbach, D., 2014. The future of sustainability science: a solutions-oriented research agenda. Sustainability science 9, 239-246.
Nicholson, C.F., 2007. Review of methods for modelling systems evolution, Discussion Paper. International Livestock Research Institute, Cornell University.
Pedercini, M., 2003. Potential contribution of existing computer-based models to comparative assessment of development options, Working Papers in System Dynamics. University of Bergen, Bergen, Norway.
Pedercini, M., 2011. Development policy analysis in Mali: sustainable growth prospects, Software Engineering and Formal Methods. Springer, pp. 447-463.
Pollitt, H., Barker, A., Barton, J., Pirgmaier, E., Polzin, C., Lutter, S., Hinterberger, F., Stocker, A., 2010. A scoping study on the macroeconomic view of sustainability, Final report for the European Commission, DG Environment. Cambridge econometrics, Cambridge, UK.
Raskin, P., Electris, C., Rosen, R.A., 2010. The Century Ahead: Searching for Sustainability. Sustainability 2, 2626-2651.
Saaty, T.L., 1990. How to make a decision: the analytic hierarchy process. European journal of operational research 48, 9-26.
Schandl, H., Hatfield-Dodds, S., Wiedmann, T., Geschke, A., Cai, Y., West, J., Newth, D., Baynes, T., Lenzen, M., Owen, A., 2015. Decoupling global environmental pressure and economic growth: scenarios for energy use, materials use and carbon emissions. Journal of Cleaner Production in press, November 2015.
Scrieciu, S., 2007. The inherent dangers of using computable general equilibrium models as a single integrated modelling framework for sustainability impact assessment. A critical note on Böhringer and Löschel (2006). Ecological Economics 60, 678-684.
Söderbaum, P., 2008. Understanding Sustainability Economics: Towards Pluralism in Economics. Earthscan, London.
Swart, R.J., Raskin, P., Robinson, J., 2004. The problem of the future: sustainability science and scenario analysis. Global environmental change 14, 137-146.
Turner, G.M., Hoffman, R., McInnis, B.C., Poldy, F., Foran, B., 2011. A tool for strategic biophysical assessment of a national economy – The Australian Stocks and Flows Framework. Environmental Modelling & Software 26, 1134-1149.
Van Beeck, N., 1999. Classification of energy models. Tilburg University and Eindhoven University of Technology.
van Vuuren, D., Hoogwijk, M., Barker, T., Riahi, K., Boeters, S., Chateau, J., Scrieciu, S., van Vliet, J., Masui, T., Blok, K., Blomen, E., Kram, T., 2009. Comparison of top-down and bottom-up estimates of sectoral and regional greenhouse gas emission reduction potentials. Energy Policy 37, 5125-5139.
Vergragt, P.J., Quist, J., 2011. Backcasting for sustainability: Introduction to the special issue. Technological forecasting and social change 78, 747-755.
Wieland, R., Gutzler, C., 2014. Environmental impact assessment based on dynamic fuzzy simulation. Environmental Modelling & Software 55, 235-241.

Wieland, V., 2012. Model comparison and robustness: A proposal for policy analysis after the financial crisis, In: Solow, R.M., Touffut, J.-P. (Eds.), What's Right with Macroeconomics? Edward Elgar, pp. 33-67.