REDEFINING MACROECONOMIC FORECASTING: A FRAMEWORK VALIDATION PROJECT

CONCEPT PAPER

WILL MAYBURY **30 DECEMBER 2025**

EXECUTIVE SUMMARY

This concept paper presents a preliminary outline of a proposed research and development initiative, Redefining Macroeconomic Forecasting (RedMaF). The project aims to deliver a validated framework for economic forecasting. It sets out the objectives, scientific rationale, and implementation strategy, while also serving as a vehicle for stakeholder engagement. It seeks to foster partnerships with individuals and organisations able to contribute specialist expertise, validation datasets, or complementary resources, and to open pathways to future funding. The content is intentionally provisional and open to refinement, allowing flexibility for collaboration and alignment with stakeholder priorities.

Despite their mathematical sophistication, current macroeconomic models lack the empirical rigour expected in disciplines that model complex systems. They cannot be systematically validated against observed outcomes, as structural errors in their mathematical formulation produce non-correctable inconsistencies. As a result, forecasts require subjective analyst adjustments, preventing independent replication. In contrast, applied sciences such as systems engineering rely on a Validation and Verification (V&V) methodology to build confidence through step-by-step empirical testing. This enables complex systems to be assembled from simpler, well-understood components, ensuring reliability, independent repeatability, and continuous improvement.

The proposed Economy Dynamics framework applies V&V principles to economic modelling. It represents the economy as a dynamic system of interacting entities, using an applied and probabilistic mathematical approach grounded in observable mechanisms. By leveraging the law of large numbers, the framework models behavioural change arising from income–expenditure imbalances, without relying on aggregate constructs such as GDP. This mechanistic approach produces forecasts that are reproducible, testable, and open to independent refinement, advancing economics towards the standards of an empirical science.

The project will be delivered by a multidisciplinary team of five experts spanning complex systems modelling, macroeconomic forecasting, economic data, V&V engineering, and software architecture. A structured three-phase programme (Concept Design, Preliminary Design, and Detailed Design) will be used to iteratively improve the framework through forecast error analysis, targeted research, and systematic refinement of models and data. The objective is not a final or perfect solution, but a foundation for cumulative scientific progress, enabling independent research teams to build upon validated components.

RedMaF mirrors the evolution of meteorology, which progressed from statistical pattern matching to applied modelling, transforming forecasting into a rigorous science. It introduces a modular, validated framework for macroeconomic forecasting that prioritises reproducibility and continuous improvement. The project will deliver open-source software with integrated validation pipelines and scalable architecture, enabling widespread testing and refinement. By replacing untestable assumptions with empirically grounded models, RedMaF establishes the basis for cumulative knowledge-building and falsifiable predictions, aligning economics with the framework of the applied sciences.

1 INTRODUCTION

Economic forecasting lacks the rigour found in disciplines that model complex systems. Although often built using sophisticated mathematics, mainstream macroeconomic models cannot be validated against observed outcomes [1]. In contrast, fields such as systems engineering demonstrate understanding through systematic validation and verification (V&V) processes [2-4]. These processes build confidence through step-by-step empirical testing, allowing complex systems to be constructed from simpler, well-understood components. By validating each element individually and then in combination, V&V enables a structured and reliable approach to modelling that avoid subjectivity and supports continuous improvement.

This project brings rigour and consistency to economic forecasting. Economy Dynamics [5] is grounded in V&V principles that model the complete economic system by using an applied and probabilistic mathematical approach. It uses the law of large numbers [6] to represent the change in entity financial behaviours under income-expenditure imbalance, based on observable mechanisms. Unlike existing forecasts, which model aggregated outputs directly (e.g. GDP), Economy Dynamics models the underlying system dynamics. The result is forecasts that are reproducible, testable, and open to independent refinement.

While ambitious, this approach has a precedent. As outlined in the Extended Synopsis, the shift from statistical pattern matching to applied modelling in meteorology is analogous [1]. This project does not aim for perfect models or datasets, but to demonstrate that applied, mechanism-based economic forecasting is capable of advancing knowledge.

The project will be delivered by a team of five experts in complex systems modelling, macroeconomic forecasting, economic data, V&V engineering, and software architecture. Through a three-phase programme, the team will improve the framework by analysing forecast errors, performing targeted research, and updating models and data accordingly. Their goal is to create the foundation for further work by independent research teams, not to provide the perfect solution. Indeed, the value of applied mathematical meteorology, was demonstrated by hand calculations [7], yet now forecasting software has more than one million lines of code.

This document sets out the scientific foundation and implementation plan for the project. It explains the limitations in the current approach, defines the V&V approach, defines the proposed forecasting framework of Economy Dynamics, and compares this to the current state-of-the-art. A proposed project description is provided at the end.

2 CURRENT LIMITATIONS

The UK Office for Budget Responsibility (OBR) uses an advanced macroeconomic model [8, 9]. First developed by the Treasury in the 1970s, it has undergone continuous development to effectively support UK fiscal policy decisions. Its framework is publicly documented, and its limitations are clearly defined [8], making it a useful illustration of the current cutting-edge.

Like other hybrid and structural models, it uses semi-empirical behavioural equations to estimate consumer spending. These equations combine economic theory, which determines the type of function and its parameters, with fitted historical data providing the values for these parameters. For example, static long-run household consumption (C) is estimated as follows [Ref 8, Equation 2.1]:

$$\text{Log}(C) = 0.96\cdot\log(RLY) + 0.04\cdot\log(NFWPE/PCE) + 0.15,$$

where RLY is real labour income, NFWPE is real financial wealth, and PCE is the consumer expenditure deflator. The theory assumes a long-term trend between the rise in household wage income and wealth income, and greater consumer spending. Though a reasonable assumption, the applicability of the selected parameters is not validated, as correlation is not causation [10]. The values of these parameters are based on correlations to data from 1972 to 2002, which was not suitable to model the global financial crash (2008), nor the economy during and after the Covid pandemic (Dec 2019 to May 2023).

The core problem in demonstrating understanding of economic behaviour lies in the presence of unexplained residuals – differences between a model's forecast and actual recent data, as shown in Figure 1. In this example, a predefined function is fitted to historical data from the UK Office for National Statistics (ONS) [11]. On the left, the function is calibrated using past data [12]; on the right, its forecast is compared against more recent data. The mismatch reveals residuals that cannot be accounted for.

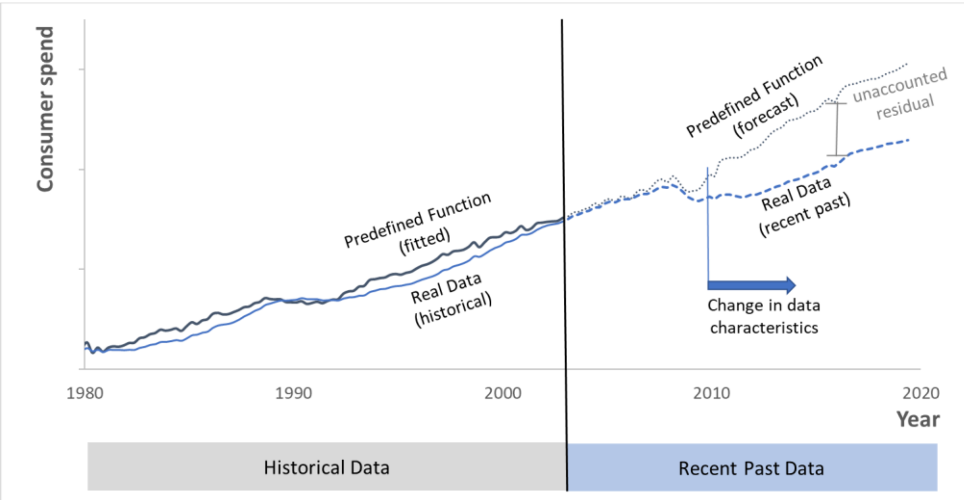

Figure 1. Illustration of unaccounted residuals in the forecast of consumer spending

The OBR recognises several limitations of its forecasting method [8]:

- Simplification: "All models are necessarily simplifications of reality".
- Unexplained behaviour: "No theory fits data perfectly".
- Human judgement: Forecasts depend on the forecaster's interpretation of residuals.
- Lacking microeconomic theory: the model lacks explicit household utility functions.
- Historical dependence: Models reflect past economic behaviour.

These limitations raise important questions: Is simplification a fundamental constraint in studying and understanding economics? Do applied science models always face unexplained behaviour? Can human judgement be made repeatable in forecasting? Is it necessary to include microeconomic theories? And does reliance on historical data preclude robust forecasting of economic behaviour after substantial structural changes? These questions are explored in the following sub-sections.

2.1 Simplification

Economic knowledge has traditionally relied on deductive reasoning, which fails to handle the complexity of real-world systems, resulting in simplistic economic theories. For example, the claim that "the value of labour is in the value of the goods" overemphasises a single mechanism – the value consumers are willing to pay – while ignoring other critical factors such as wage competition, sectoral cost differences, regional cost of living, and business profit requirements. This limitation in a deductive approach is avoided by modelling the dynamics of the system.

The simplification in economic theories also necessitates the use of idealised assumptions, such as perfect competition and rational consumers, which are not analogous to those of science. In science, simplified models like the ideal gas law are constrained to well-defined 'idealised' conditions under which they are empirically valid. If those conditions are not met, different methods are used. In economics, by contrast, idealised assumptions do not occur in reality and do not constrain the model's suitability [13, 14].

2.2 Unexplained Behaviour

The current use of semi-empirical models fitted to aggregated historical data may introduce several uncorrectable errors. These can be broadly categorised into data fitting, variable selection and aggregation errors, discussed by Maybury (2020, Paper 00 [5]). In summary:

Data fitting errors arise from a training dataset that is too narrow (Figure 2 [15]), or when an inappropriate function type is used to represent the data, producing unaccountable residuals.

Variable selection errors occur when important explanatory variables are omitted, or where the relationship between cause and effect is not validated by a separate methodology. The result is that relationships between variables that appear statistically significant may be misleading [16].

Aggregated data errors occur when functions are applied to aggregated and not disaggregated inputs. For example, summing regional economic inputs (e.g., income) before calculating the output (e.g., consumption) does not produce the same results as modelling each region individually and then summing the outputs. This is illustrated in Figure 3.

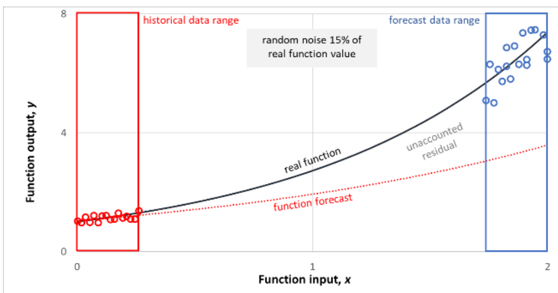

Figure 2. Illustration of sample range error

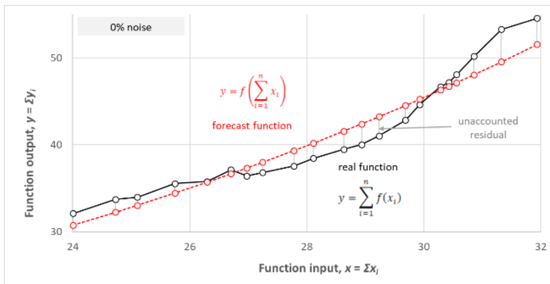

Figure 3. Illustration of aggregation errors

Semi-empirical models cannot reliably produce accurate forecasts when extrapolated or aggregated. In contrast, Economy Dynamics avoids this by designing models to operate strictly within known data ranges (i.e., interpolation only) and by using a disaggregated methodology (i.e., household subgroup, industrial sectors and economic region). For example, because households have experienced a wide range of financial conditions in the past, their behavioural patterns (captured in disaggregated datasets) enable forecasting through interpolation. This enables robust simulation of new policies even after substantial economic structural changes.

2.3 *HUMAN JUDGMENT*

Errors in current macroeconomic models that lack meaningful, quantifiable corrections often force analysts to rely on individual judgment to align model outputs with historical data, introducing errors due to the subjectivity of the analysts.

In contrast, validation is a cornerstone of applied sciences, offering a systematic way to assess and compare models based on how well they explain observed behaviour under specific conditions. It not only enables objective evaluation but also enables consistency and credibility in model use. For example, in engineering, software models are routinely validated against established datasets, demonstrating both the model's reliability and the analyst's competence. Best-practice methodologies evolve through continuous feedback and collaboration to achieve consistent, high quality validation evidence throughout the community of analysts.

2.4 *LACKING MICROECONOMIC THEORIES*

As outlined in the Extended Synopsis [1], the disconnect between theoretical and empirical coherence necessitates the rejection of theories under standard scientific methodology [e.g., 17, 18]. Consequently, excluding unvalidated theories is not a shortcoming but a necessary step in advancing our understanding.

2.5 *USING HISTORICAL DATA*

Forecasting household and business behaviour doesn't need to depend on projecting correlations into the future. Instead, it can be approached through interpolation, much like weather forecasting, where the primary source of error stems from the butterfly effect [19], a theory that also influences economic systems. This shift requires a mechanistic framework: models built on validated empirical observations, incorporating probabilistic elements when needed (for instance, surrogate models to represent behavioural responses). Forecasts will still contain errors, these can be quantified and traced to specific, correctable sources, eliminating the reliance on subjective analyst judgment.

3 VALIDATION AND VERIFICATION (V&V) METHODOLOGY

The methodology used in this project is based on decades of engineering best practice [e.g., 2-4] and follows a structured, four-step process: 1) Operational Requirements Definition, 2) Modular Decomposition, 3) Development and Integration, and 4) Testing. This process is commonly represented by the V-diagram (Figure 4), where the left arm represents system decomposition focusing on requirements management and test planning. The right side continues with development, software construction, and testing. The whole process is repeated to increase software maturity through each of the three phases: Concept Design, Preliminary Design, and Detailed Design.

Step 1: Operational Requirements Definition. Projects begin by defining what the software (system) must achieve. The output is the Operational Level requirements, such as capturing the Central Bank's requirements for economic forecasting tools. For example, this will include the customer goals - forecasting the effect of interest rate changes on inflation, and functional needs – modelling changes in consumer spending due to changes in interest rates.

Step 2: Modular Decomposition. The forecasting software is structured into three hierarchical levels. Requirements are broken down in a top-down manner across the different levels. Tests are planned for every requirement level to define the assessment of fulfilment:

- Software Level – simulates whole-economy behaviour under any scenario.
- Model Level – integrates behavioural modules for households and businesses, financial calculations, and price-setting mechanisms.
- Method Level – defines functions and algorithms that drive model behaviour, such as those accounting for interest rate changes or tax adjustments in the calculation of household and business financial circumstances.

Step 3: Development and Integration. The development phase begins with defining the methods – mathematical functions, then the models – interaction between methods and data handling, and finally the software framework – the architecture of the interaction between models and data handling. After being defined, each stage is coded and verified against its respective level requirements before moving to the next level, i.e., methods => models => software. Finally, it must be demonstrated that the software produces the forecasts that explain observed behaviour. Once this is all completed the software can be tested in the operational environment in Step 4.

Step 4: Testing and Validation. The software is validated by testing it under operational conditions to ensure it meets the original operational-level requirements defined in Step 1. This confirms that the system performs as intended in its real-world context. For example, it explains observed outcomes, such as historical inflation under real-world conditions.

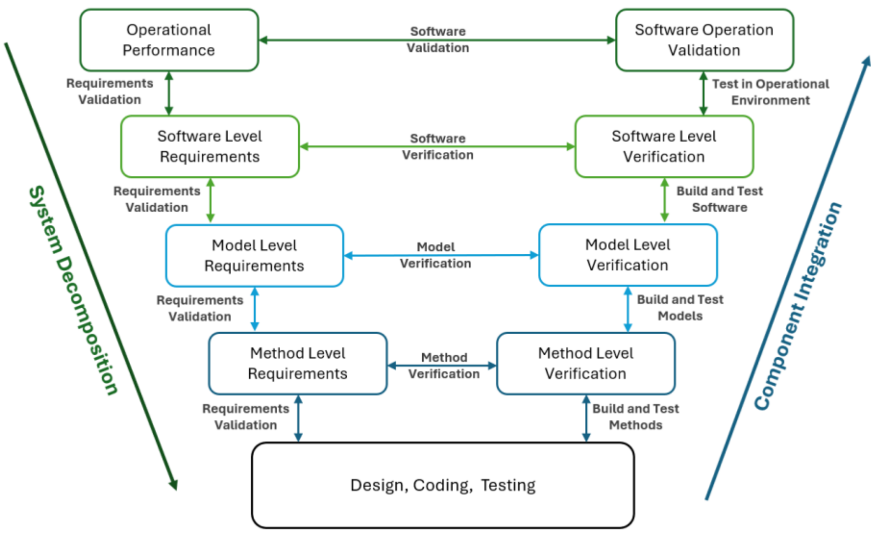

Figure 4. V-Diagram

In systems engineering, the terms validation and verification are typically defined as follows [4]:

Validation of Operational Performance means checking that the system works as intended in its real-world environment through testing in the operational environment, and not the development environment (see the top row of Figure 4). While systems engineers focus on validating the whole system, technical specialists, who develop specific methods or models, also need to follow the V-diagram process. These validation results are then documented and used by systems engineers to help confirm that everything works together correctly (see Verification of Component Integration).

Verification of Component Integration confirms that the requirements defined at the decomposition stage have been fulfilled across methods, models, and software. While validation is carried out by technical specialists at each level, verification assures that these results collectively demonstrate compliance with the specified requirements. This stage does not involve re-running operational tests (see *Validation of Operational Performance*), but rather reviewing the evidence that each component behaves as intended and integrates consistently.

4 FRAMEWORK STRUCTURE AND RATIONALE

The Economy Dynamics software simulates the economy as a system composed of disaggregated households and businesses, distributed across sectors and regions. Transactions are modelled as probabilistic events, determined by financial context and constraints. Goods and services are grouped hierarchically (e.g., food > cereals > rice), and households and businesses are grouped by similar financial and demographic characteristics.

The software can aggregate or disaggregate each domain, allowing high resolution in critical areas (e.g., housing in London). The software architecture ensures this flexibility while maintaining the integrity of forecasts.

The approach is mechanistic, representing observable processes, where a mechanism is any feature or process that affects the probability of a transaction occurring, such as taxation, wages, prices, and even laws. Mechanisms are implemented as methods, combined into models (e.g., household finance), and integrated into software modules.

This modular architecture allows for validation at multiple levels. For example, a method calculating income tax can be tested independently before being integrated into the household model. The household model can then be validated against national income data.

4.1 Requirements Driven

The Economy Dynamics forecasting software is a requirements-driven approach, following the V&V methodology (Section 3). A small selection of the current (functional) requirements is shown in Table 1, for reference. Note that the word ‘shall’ has a formal meaning, indicating that the requirement must be fulfilled within the project.

Level	Description
Operational Level	Software shall produce testable forecast metrics for monetary policy analysis.
Software Level	Software shall simulate the change in features of an economy over time.

Level	Description
Software Level	Consumers shall include all entities that consume goods, services or assets.
Model Level	Models shall include critical verifiable processes that affect the price of supplies.
Model Level	Models shall include critical verifiable processes that affect the wages.
Model Level	Models shall include critical verifiable processes that affect the house rental cost.
Model Level	Models shall link income with expenditure to determine financial circumstances.
Model Level	Decision making shall model behavioural change due to financial circumstances.
Method Level	Methods shall represent the spatial effects on the price of goods and services.
Method Level	Methods shall represent the temporal effects on the price of goods and services.
Method Level	Retailer pricing shall include the supply chain effects on the cost of production.
Method Level	Methods shall represent tax effects on entity financial circumstances.
Method Level	Methods shall represent interest rate effects on entity financial circumstances.
Method Level	Methods shall represent debt repayment on entity future financial circumstances.

Table 1. Sample of the Software Requirements for the Economy Dynamics Framework

4.2 Economy Dynamics Framework

The framework methodology is outlined by Maybury, 2020, Paper 00 [5]. The key aspects include:

Discretisation: The division of data into subgroups or elements, also called disaggregation. Households are grouped into demographic subcategories that share similar financial behaviours. Businesses are classified into subgroups based on industry, size, age, and business practices, which may reflect differences in quality, sustainability, or ethical approaches. Geographic regions are divided into economic units with internal homogeneity, with separate classifications for households and businesses. This structure ensures that variations in costs, tax rates, and income levels are accurately represented.

Validation: The simulation advances in discrete timesteps, updating the financial circumstances of entities, supply prices, and financial and environmental output measures. This enables the testing of dynamic trends over time, such as inflation or the effects of policy changes. Model features will be validated against observed economic and environmental indicators, including GDP, inflation, employment levels, and the carbon footprint of consumption. To ensure reliability, an automated pipeline (common in software development) will be used so that ongoing improvements to the code do not compromise previously validated results.

Scalability: The computational framework is designed for modularity and scalability. Analysts can control the simulation complexity by specifying the level of discretisation of households, businesses, regions and supplies. High-resolution simulations are feasible where data, time and high-performance computational facilities are available. This approach allows for valid simplification within appropriate constraints of prescribed, tested assumptions.

4.3 Economy Dynamics Models

Economy Dynamics simulates the entire economic system by modelling the observable mechanisms, which are represented by the following models: Household Income Forcing (response to income-expenditure imbalance) [20]; Macroeconomic Primitive [21]; Retailer Price-Setting [22]; and the models in preparation, Life-Cycle Events [23], Business Income Forcing [24], Land Pricing [25], and Wage Rates for Labour [26]. The features of these models are as follows:

Behavioural Modelling: Consumer and business decision-making is derived from population data using surrogate models (e.g. regression, machine learning), trained to predict financial behaviour under various income and expenditure conditions. These models determine spending and saving behaviours, and investment and borrowing decisions. The mathematical framework ensures that surrogate models only interpolate correlations and do not extrapolate them. For the approach to modelling households see Maybury 2020, Paper 01 [20].

Complete System: The macroeconomic primitive is an equation that describes the income and expenditure of all entities. In its expanded form, it defines a system of financial transactions where every expenditure has a corresponding income, interlinking all transactions. This includes the effects of trading between economic regions and the dynamic influence of time, such as the lag between purchasing input goods from one region, paying wages in another, producing goods, selling them and receiving payment. The equation explicitly models regional fiscal and monetary policies. For the method definition see Maybury 2020, Paper 02 [21]

Price Resolution: Prices for goods, services, labour, and land are determined by the seller/retailer, based on information available. For supplies the model assesses the cost of production (including time and location-dependent inputs), the financial circumstances of suppliers and consumers, and the regional availability and demand. For the definition of the price of supplies, see Maybury 2025, Paper 03 [22].

Lifecycle Modelling: Households and businesses undergo life cycle transitions. For households, these include births, education & employment transitions, cohabitation, migration and death. All of these affect household financial status and decision-making. These transitions are modelled probabilistically and introduce structural changes to the simulated economy over time. This enables demographic shifts to be captured in the forecast. The approach is summarised in the framework overview by Maybury, 2020, Paper 00 [5].

5 STATE-OF-THE-ART COMPARISON

This section provides a summary of the comparative differences between four widely used forecasting frameworks and Economy Dynamics. The key differences are shown in Table 2.

Structural models do not represent observable mechanisms [e.g., 27-32]. These include fully theoretical coherent methodologies, such as the Dynamic Stochastic General Equilibrium (DSGE) and partial theoretical coherent models, such as large-scale forecasting models.

Hybrid models incorporate some mechanisms such as price = cost + margin [e.g., 33,34], including Stock-Flow Consistent models (SFC). They introduce partial realism through selected observed mechanisms that appear to give outcomes consistent to unvalidated theories.

Statistical models lack theory or mechanistic representation of the system. They forecast output variables directly from correlated past trends [e.g., 35-37], such as Vector Autoregression (VAR).

Agent based models (ABM) represent households and firms as instantiated computational objects [e.g., 38,39], which are copies of behavioural models with independent datasets. This allows the simulation to represent the entire interaction between households and firms, with any number of individual entities. Their interaction can represent the evolution of a closed system, such as a housing market. They cannot be meaningfully scaled to whole-economy modelling. **Economy Dynamics** (ED) a new forecasting framework that has never been implemented previously: it models the complete economic system using observable mechanisms, disaggregated data, and a modular architecture grounded by a validation and verification (V&V) methodology. Unlike existing approaches, it does not rely on statistical extrapolation or untested assumptions. Instead, it represents how real households and businesses behave, using surrogate models trained at the level of economic regions, and population subgroups. Crucially, it is the first framework designed to be systematically validated and progressively refined, enabling economics to transition from speculative theory to a reproducible, testable applied science.

<table>
<tr><th>Features</th><th>Structural (DSGE)</th><th>Hybrid (SFC)</th><th>Statistical (VAR)</th><th>Agent-Based (ABM)</th><th>Economy Dynamics (ED)</th></tr>
<tr><td rowspan="2">Framework can be validated</td><td>✗</td><td>✗</td><td>✗</td><td>◐</td><td>✓</td></tr>
<tr><td colspan="5">ABMs for small scale simulations; ED at all levels of an economy.</td></tr>
<tr><td rowspan="2">Avoids statistical extrapolation</td><td>✗</td><td>✗</td><td>✗</td><td>✓</td><td>✓</td></tr>
<tr><td colspan="5">DSGE, SFC and VAR rely on extrapolation; ED uses surrogate modelling of behaviour under varying financial circumstances</td></tr>
<tr><td rowspan="2">Models the system behaviour and not just outputs</td><td>✗</td><td>◐</td><td>✗</td><td>✓</td><td>✓</td></tr>
<tr><td colspan="5">SFC models partially represent the system but are biased by unvalidated theories; ABM and ED model mechanisms</td></tr>
<tr><td rowspan="2">Accounts for all crucial information i.e., variables, functions, and models.</td><td>✗</td><td>✗</td><td>✗</td><td>✗</td><td>✓</td></tr>
<tr><td colspan="5">ED by implementing the V&V methodology ensures that the system models are (sufficiently) complete.</td></tr>
<tr><td rowspan="2">Disaggregated datasets into entity subgroups, economic regions, industrial sectors, …</td><td>✗</td><td>✗</td><td>✗</td><td>✗</td><td>✓</td></tr>
<tr><td colspan="5">Only ED models regions independently and allows validation per region and by sector.</td></tr>
<tr><td rowspan="2">Observable mechanisms</td><td>✗</td><td>◐</td><td>✗</td><td>✓</td><td>✓</td></tr>
<tr><td colspan="5">SFCs assumes pricing rules that align with unvalidated theories; ED models business production costs and competition based on all observed mechanisms.</td></tr>
<tr><td rowspan="2">Modular architecture</td><td>✗</td><td>✗</td><td>✗</td><td>✗</td><td>✓</td></tr>
<tr><td colspan="5">ED supports model (and method) substitution as part of the V&V enabling independent piecewise research of each component.</td></tr>
</table>

Legend: ✓ = Fully supports feature; ◐ = Partial / constrained support; ✗ = Not supported

Table 2. Comparative Table of Forecasting Models

Economy Dynamics (ED) introduces a forecasting framework that has never been implemented before: it models the complete economic system using observable mechanisms, disaggregated data, and a modular architecture grounded by a validation and verification (V&V) methodology. Unlike existing approaches, it does not rely on statistical extrapolation or untested assumptions. Instead, it represents how real households and businesses behave, using surrogate models trained at the level of economic regions, and population subgroups. Crucially, it is the first framework designed to be systematically validated and progressively refined, enabling economics to transition from abstract theory to a reproducible, testable applied science.

6 PROJECT DESCRIPTION

6.1 WORK PLAN AND METHODOLOGY

The project is split into five work packages (WPs). Two are for managing the project (WP0) and the software (WP4). The other three are for the advancement of science and are structured around three phases: Concept Design (WP1), Preliminary Design (WP2), and Detailed Design (WP3). Each phase follows a consistent sequence of six tasks (Figure 5): Framework Definition, Data Discretisation, UK Database Curation, UK Economy Model, Build Software, and Validation. This staged process ensures rigorous Verification and Validation (V&V) for iterative error reduction through method, model and software refinement. The work packages are summarised below, with a Gantt chart presented in Figure 6.

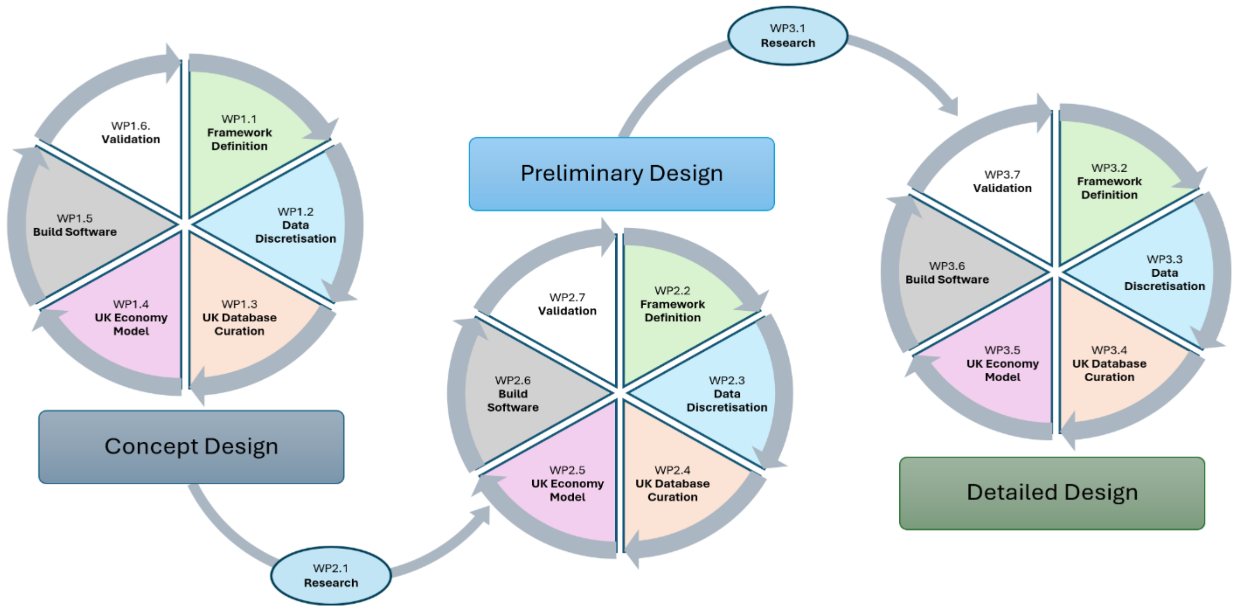

Figure 5. Task Flow Through the Concept, Preliminary and Detailed Design Phases

6.2 WP0 PROJECT MANAGEMENT AND DISSEMINATION

Recruitment & Resourcing: Hiring a forecasting model developer, an economic data analyst, and a systems software developer. If recruitment is unsuccessful, subcontracting and consultancy options will be pursued. **End User Engagement:** Regular outreach to academic centres, public institutions, and consultancies to inform, invite collaboration, and tailor development to real-world needs. **Dissemination:** Research outputs will be presented at conferences, published in journals, and a monograph will be prepared by the end of the project. **Monitoring & Reporting:** Weekly technical meetings, Monthly technical reviews, Quarterly project reviews, ERC reporting, and frequent risk tracking, as applied to other R&D projects.

6.3 WP1 Concept Design / WP2 Preliminary Design / WP3 Detailed Design

Each phase completes the same major technical tasks, with updates in Phases 2 and 3 guided by validation results from the previous phase:

1. Framework Definition – WP1.1 / WP2.2 / WP3.2

Development of the modular Economy Dynamics framework integrating validated surrogate and applied mathematical models: 1) household, business, and retailer decision-making; 2) system-wide financial interdependencies; 3) normalised pricing models for land and labour; 4) validation methodology and computational architecture specification.

2. Data Discretisation – WP1.2 / WP2.3 / WP3.3

Subdivision of the economy into probabilistic units. This structure supports both fine and coarse-grained simulations using a single data architecture. The subgroups include:

- Household types (income, demographics, job categories).
- Business categories (sectors, input/output profiles).
- Supplies (goods/services hierarchies) and assets (land usage/financial categories).

3. UK Data Curation – WP1.3 / WP2.4 / WP3.4

Curation and processing of historical UK data to support: Surrogate model training; Empirical validation of methods, models and software; Custom datasets reflecting UK-specific tax, welfare, and inflation factors.

4. UK Economy Model – WP1.4 / WP2.5 / WP3.5

Adapting models for the UK economy. Develop the applied mathematics models for the UK by the appropriate expansion of the macroeconomic primitive equations. In addition, develop wage and land-pricing models using a engineering-type normalisation approach, whereby values are expressed relative to other calculated financial terms within the framework, including the cost-of-living and the accumulation of wealth. This ensures that prices emerge from defined relationships rather than being imposed externally.

5. Build Software – WP1.5 / WP2.6 / WP3.6

Modular software development including functional code libraries for each model, unit testing and simulation control; flexible data handling to manage varying levels of simulation fidelity; and a central interface allowing forecasters to easily configure simulations.

This task includes the validation of methods (functions) and models (sets of integrated functions), as well as defining corresponding unit tests within the software validation pipeline. Once validated and tested, all future software updates will be subject to these unit tests to ensure new developments do not compromise the integrity of previously validated methods, models, or the complete software. The validation tests will be established to satisfy the requirements identified in Work Package 4.1 (see) and an example is shown in Table 1.

6. Validation – WP1.6 / WP2.7 / WP3.7

Validation of the software operation is shown in the upper right of Figure 4. This defines the final validation cases to run, such as GDP, inflation, unemployment and consumer carbon footprint, and which observed historical trends they are to be compared to. The operational validation tests that the software satisfies the requirements identified in Work Package 4.1 (see Figure 6) for operational performance (see Table 1 for example requirements).

7. Research – WP2.1 / WP3.1

At the end of the Concept and Preliminary Design phases, errors in the validation results will be researched. The output of the research is to propose revisions to the methods, models and software approach, to overcome shortfalls in the validation. A review of the requirements to confirm that they are correct and complete. The results feed into the framework definition and follow-on tasks in the next phase.

6.4 WP4 Software Management

Software management processes: prepare and update requirements and QA documentation across all phases; create version control with a continuous integration pipeline; identify parallel processing capabilities for scalable execution; develop user interface and developer documentation to support uptake by other researchers; and release code as open source, with active community management.

6.5 Gantt Chart and Milestones

The Gantt Chart below (Figure 6) shows the distribution of tasks across the project's timeline.

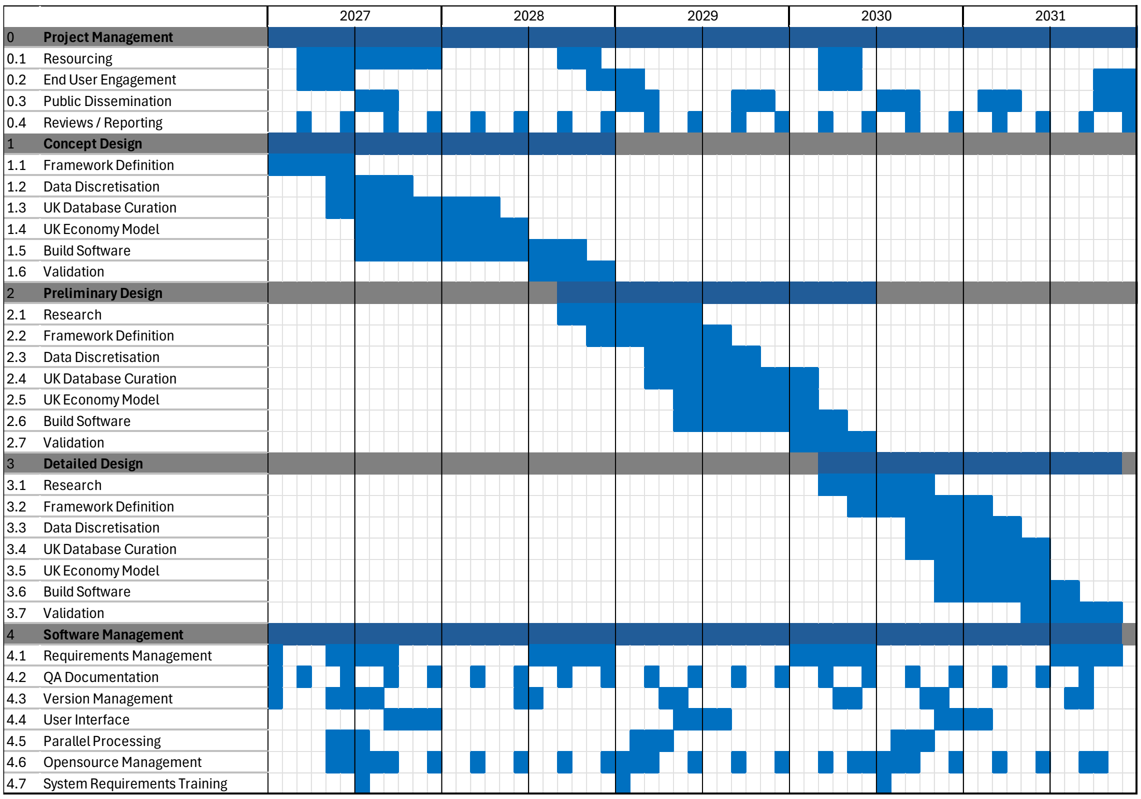

Figure 6. Project Gantt Chart

The key milestones for recruitment, set-up of the open-source software framework, end of phase software build releases, and the scientific dissemination schedule are listed below:

Milestone 0: Team Established (Month 6). Recruitment of all staff, the signing of employment contracts, and the agreement of official start dates.

Milestone 1: Open Software Framework (Month 8). From Month 8 onward, the software will be managed through an open-source version control platform (such as GitHub). All future development will be conducted on this platform, ensuring that the most up-to-date version of the software is publicly accessible.

Milestone 2: Concept Phase Complete (Work Package 1, Month 24). The phase begins by defining the software framework, followed by specifying the data discretisation (disaggregation). This is followed by creating the initial UK database and UK-specific models, developing the concept prototype software, and finally performing the V&V testing and requirements reviews (Figure 5, left).

Milestone 3: Preliminary Phase Complete (Work Package 2, Month 42). Errors in the validation of the Concept Software will be quantified to establish areas of shortcomings. Research to improve the framework, models and data will be documented. The framework definition updated, data discretisation refined, UK database improved, UK specific models revised, preliminary prototype software built, and the Preliminary Phase V&V processes completed (Figure 5, middle).

Milestone 4: Detailed Phase Complete (Work Package 3, Month 59). Errors in the validation of the Preliminary Software will be quantified to establish areas of shortcomings. Research to improve the framework, models and data will be documented. The framework definition updated, data discretisation refined, UK database improved, UK specific models revised, detailed prototype software built, and the Detailed Phase V&V completed (Figure 5 right).

Milestone 5: Scientific Dissemination (Ongoing). A key focus throughout the project will be the scientific dissemination. At least two peer-reviewed publications will be submitted to open-access journals at the end of each phase (Months 24, 42, and 59) and presentations will be submitted to two conferences. During the Preliminary and Detailed phases, free-to-attend workshops will be held. A comprehensive monograph will be published at the project conclusion.

6.6 Risks and Opportunities

The following provides a summary of the key risks. All risks will be tracked regularly during the project using a live risk register. New risks and their mitigation will be identified continuously and reviewed formally each month. Risks arising will be assessed and addressed at weekly meetings, and in practice informally between the relevant tech staff and principal investigator as they arise.

Data Availability (High Risk/Medium Impact). If specific input datasets are unavailable, the project will pursue a tiered mitigation strategy. This includes identifying equivalent datasets from other economies, conducting targeted small-scale surveys, collaborating with relevant academic or institutional partners [e.g., 40,41], or constructing well-documented synthetic datasets based on transparent, testable assumptions. These assumptions will be published to invite external feedback. While the risk of missing data is real, it is not a project-ending constraint. It is expected that at least some required datasets will be incomplete or unavailable in the desired form. In such cases, initial assumptions can be revised in Phases 2 and 3 (WP2 and WP3) as feedback is received and understanding improves.

Note that public statistical bodies (e.g., the UK ONS) will not collect data in the required structure until demanded by the forecasting community. However, as demonstrated in the early development of weather forecasting [7], working with incomplete data is sufficient to validate the approach and prove its potential. This, in turn will lead to the demand for the correctly structured data.

Data Disaggregation (High/Low). Techniques such as agent-based modelling (ABM) will be applied. Note: the aim is to advance knowledge, not deliver a commercial software toolset.

Task Overrun (High/Low). A short-term "make-do" approach will be applied, with recovery planned for the next phase.

Model Quality (High/Low). Early models will be treated as hypotheses and iteratively refined. Note that the concept of weather forecasting was demonstrated with just 12 equations.

Model Approach (Medium/Medium). If the engineering normalisation method for generalising the calculation of land prices or labour wages based on other variables such as investment wealth or the cost-of-living cannot be achieved, a surrogate modelling approach will be used.

Unproductive Staff (Low/High). The project will prioritise hiring senior-level professionals with over 15 years of proven achievement. A probation period will be implemented.

No Suitable Applicants (Low/Medium). Top-tier salaries will be offered and subcontracting or service providers such as NISER may be used while recruitment continues.

High-Performance Computing (Low/Medium). Usage will be monitored, and additional funding will be sought if necessary if the project's computational demand has been underestimated. If usage has been overestimated, then this represents an opportunity to reallocate the funds.

Opportunity: AI will support software development, generation of the quality assurance documentation and software unit testing. It will be treated as a knowledgeable, mistake prone assistant. All AI outputs will be comprehensively reviewed by a technical expert.

6.7 *Justification of Staff Roles*

Principal Investigator (PI): Responsible for overall management and technical leadership of the project, ensuring delivery in line with ERC objectives. The PI provides the scientific vision and architectural oversight to guide model and software development, while coordinating technical work, progress reporting, and resource allocation. A further role is team development: recruiting staff with transferable skills, supporting their adaptation to the novel methodology, and building new technical competencies. The PI also ensures outputs are effectively communicated to academic communities and external stakeholders, including research institutes, consultancies, and policy organisations, to maximise the project's broader impact on the field of economics.

Macroeconomic Model Developer (staff): Manage model development, contribute to the software development, run forecasts, write papers, and research improvements. Main challenge: develop normalised engineering-type models of pricing (e.g. housing, and labour) and develop the applied mathematics models for the UK (e.g. the macroeconomic primitive).

Economics Data/ABM Analyst (staff): Manage datasets, contribute to the software development, supporting forecasting, writing papers, and research improvements. Main challenge: curate large-scale population and economic datasets for the UK, disaggregated by household, region, and sector. This requires the development of methods to estimate probabilistic subgroup dynamics not available in the existing datasets.

Software Developer / Systems Architect (staff): Manage the software development, support forecasting, write papers (software framework), and research improvements. Main challenge: design and implement a scalable data architecture, enabling flexible aggregation-disaggregation and efficient processing. Requires expertise in parallel computing.

Communications Assistant (staff): Support outreach to the scientific community, project visibility, submission to publication. This role ensures stakeholder engagement and public access to the project's outputs including reports, datasets and software.

Software Systems Engineer (Subcontract): A software requirements specialist will deliver initial training and provide reviews to ensure compliance with the formal V&V methodology.

6.8 Publications and Exploitation

The project's publication strategy prioritises open access and reproducibility, with peer-reviewed papers published annually and a monograph produced at the project's conclusion. To ensure methodological transparency and foster broader uptake, all models, data structures, and validation methods will be openly documented and shared where possible.

7 PROJECT OBJECTIVES AND IMPACT

The primary objective of this project is to address the lack of validated knowledge in the field of economic forecasting, which will have spill-over benefits in both micro and macroeconomic research. This will be achieved by developing a novel forecasting framework (Economy Dynamics) that replaces untestable economic assumptions with a modular system of empirically grounded, testable models. Mirroring the Verification and Validation (V&V) methodology from engineering, the framework will decompose macroeconomic behaviour into a series of observable mechanisms. Each model will be independently verified and validated against historical datasets, enabling cumulative scientific progress and transparent assessment of alternative theories. This approach will remove reliance on unvalidated assumptions such as utility functions and statistical extrapolation of GDP trends, which currently undermine the reliability and interpretability of economic forecasts.

A secondary objective is to deliver a fully documented, open-source software platform that allows researchers and policymakers to construct, test, and extend economy simulations using the Economy Dynamics framework. This will include a configurable database structure to enable disaggregation by sector, region, and entity type; automated validation pipelines; and a simulation architecture that can scale from basic national forecasts to complex interregional economic systems. The framework will provide a foundation for reproducible economic science, enabling independent teams to replicate results, compare theories, and integrate new mechanisms as knowledge evolves – bringing macroeconomic forecasting into alignment with the standards of the applied sciences.

The scientific impact of this project is that it redefines the foundations of macroeconomics by introducing a validated, verifiable framework that enables falsifiable predictions. This is a crucial concept in the philosophy of science that I want to apply to economic behaviour. By replacing abstract assumptions with modular, testable mechanisms grounded in observable data, the Economy Dynamics framework unlocks new horizons for empirical research: enabling theory comparison and eventually, unifying the field of economics and deriving new collaborative subfields. It creates the infrastructure for cumulative knowledge-building, providing scientific rigour to the study of complex issues from income distribution to environmental impact.

8 REFERENCES AND ENDNOTES

1. Maybury, W. J. (2025). Redefining Macroeconomic Forecasting: validation of a new framework. Axsym Engineering Publications, Economy Dynamics White Paper 01: http://www.axsym-engineering.com/publications/EconomyDynamics_WhitePaper01.pdf
2. Thacker, B., Doebling, S. W., Hemez, F. M., Anderson, M. C., Pepin, J. E., and E. A. Rodriguez (2004). *Concepts of model verification and validation*. Los Alamos National Lab Report, LA-14167. Los Alamos, NM, United States. https://www.osti.gov/biblio/835920
3. Rook, P. (1986). *Controlling Software Projects*. Software Engineering Journal 1, pp. 7-16. https://doi.org/10.1049/sej.1986.0003
4. An example of an industry standard for systems engineering best practices: INCOSE (2015). *Systems Engineering Handbook*.
5. Maybury, W. (2020 [Paper 00]). Towards a comprehensive macroeconomic simulation capability: Economy Dynamics. Axsym Engineering Publications: https://www.axsym-engineering.com/publications/EconomyDynamics_Paper00.pdf
6. Bolthausen, E. and M. V. Wuthrich. (2013). *Bernoulli's Law of Large Numbers*. Astin Bulletin 43(2), 73-79. doi: 10.1017/asb.2013.11
7. Richardson, L. F. (1922). *Weather Prediction by Numerical Process*. Cambridge University Press.
8. OBR (2013). *The macroeconomic model*. Office for Budget Responsibility, Briefing paper No. 5: https://obr.uk/docs/dlm_uploads/Final_Model_Documentation.pdf
9. OBR (2021). *OBR macroeconomic model*. OBR Reports: https://obr.uk/forecasts-in-depth/obr-macroeconomic-model/
10. Aldrich, J. (1995). *Correlations Genuine and Spurious in Pearson and Yule*. Statistical Science. 10 (4): 364–376. doi:10.1214/ss/1177009870. JSTOR 2246135
11. ONS, 2019. Household final consumption expenditure: National concept CVM SA - £m. Office for National Statistics (UK): Economic, social and population statistics publication: https://www.ons.gov.uk/economy/nationalaccounts/satelliteaccounts/timeseries/abjr/ukea
12. The function used in Figure 1 has been created by the author to be representative of the OBR [8] methodology, for expedience it is not the exact function. The unaccounted residuals are representative of the effects of a change in economic structure.

13. Kapeller, J. & S. Pühringer (2010). The internal consistency of perfect competition. Journal of Philosophical Economics, Vol III, 2 (1). http://dx.doi.org/10.46298/jpe.10598
14. Raworth, K. (2017). *Doughnut Economics: Seven Ways to Think Like a 21st-Century Economist*. Chelsea Green Publishing.
15. The dataset is constructed to illustrate sample range error. Historical data spans the range 0≤ x <0.30 (red open circles) and is used to fit an exponential function for forecasting y (red dotted line). Following a structural shift the relevant data range becomes 2.7 ≤ x < 3.0, resulting in unaccountable residuals between the forecast and reality.
16. Aldrich, J. (1995). *Correlations Genuine and Spurious in Pearson and Yule*. Statistical Science. 10 (4): 364–376. doi:10.1214/ss/1177009870. JSTOR 2246135
17. Popper, K. (1934). *The Logic of Scientific Discovery*. 1959 translation by the author, reprinted by Routledge Classics, London and New York (2002)
18. Feynman, R. (1964). *Scientific Method*. Lectures to undergraduates at the California Institute of Technology: (Caltech). https://www.youtube.com/watch?v=EYPapE-3FRw&t=100s
19. Gleick, J. (1991). *Chaos: The making of a new science*. Cardinal, Sphere Books Ltd, London.
20. Maybury, W. J., 2020 [Paper 01]. Modelling decision-making from population data: Household Income Forcing. Axsym Engineering Publications, Economy Dynamics Paper 01: https://www.axsym-engineering.com/publications/EconomyDynamics_Paper01.pdf
21. Maybury, W. J., 2020 [Paper 02]. The economy in one equation: The Macroeconomic Primitive. Axsym Engineering Publications, Economy Dynamics Paper 02: https://www.axsym-engineering.com/publications/EconomyDynamics_Paper02.pdf
22. Maybury, W. J., 2025 [Paper 03]. Retailer decision-making: The price of supplies. Axsym Engineering Publications, Economy Dynamics Paper 03: https://www.axsym-engineering.com/publications/EconomyDynamics_Paper03.pdf
23. Maybury, W. J., *in progress* [Paper 04]. *Consumer life cycles: Microeconomic dynamic change*. Axsym Engineering Publications, Economy Dynamics Paper 04.
24. Maybury, W. J., *in progress* [Paper 05]. *Modelling decision-making from population data: Business Income Forcing*. Axsym Engineering Publications, Economy Dynamics Paper 05.
25. Maybury, W. J., *in progress* [Paper 06]. *The price of land*. Axsym Engineering Publications, Economy Dynamics Paper 04.
26. Maybury, W. J., *in progress* [Paper 07]. *The wage rate for labour*. Axsym Engineering Publications, Economy Dynamics Paper 07.
27. Correa-Jimenez, A., Davies, R., and R. Siddique, 2024. *Macroeconomic Modelling: A Review*. Economics Observatory and LSE Growth Co-Lab.
28. NIESR. (2023). *The leading global macroeconomic model*. NIESR, Macroeconomic Model: https://www.niesr.ac.uk/nigem-macroeconomic-model.
29. Burgess, S., Fernandez-Corugedo, E., Groth, C., Harrison, R., Monti, F., Theodoridis, K., and Waldron, M. (2013) *The Bank of England's forecasting platform: COMPASS, MAPS, EASE and the suite of models*. Working paper no. 471. Bank of England.
30. Laxton, D. (2008). *Getting to Know the Global Economy Model and its Philosophy*. IMF Staff Papers Vol 55 No. 2.
31. Vines, D., & Wills, S. (2018). *The Rebuilding Macroeconomic Theory Project*. Oxford Review of Economic Policy.
32. Kydland, F. and Prescott, E. (1982). *Time to Build and Aggregate Fluctuations*. Econometrica 50(6).

33. Godley, W. & Lavoie, M. (2007). *Monetary Economics: An Integrated Approach to Credit, Money, Income, Production and Wealth*. Palgrave McMillan, Basingstoke, Hampshire.
34. Caverzasi, E., and A. Godin. 2015. *Post-Keynesian Stock-Flow Consistent Modelling: A Survey*. Cambridge Journal of Economics 39 (1): 157–187. DOI: 10.1093/cje/beu021
35. Blanchard, O. J., and D. Quah. (1989). *The Dynamic Effects of Aggregate Demand and Supply Disturbances*. American Economic Review 79 (4): 655–673.
36. Sims, C. A. (1980). *Macroeconomics and Reality*. Econometrica 48 (1): 1–48. DOI: 10.2307/1912017.
37. Box, G., Jenkins, G., Reinsel, G., and G. Ljung (1970). *Time Series Analysis - Forecasting and Control*. John Wiley & Sons, 5th Edition (2016).
38. Schelling, T. C. (1971). Dynamic models of segregation. The Journal of Mathematical Sociology 1 (2). https://doi.org/10.1080/0022250X.1971.9989794.
39. Axtell, R. L., and J. D. Farmer. (2025). *Agent-Based Modeling in Economics and Finance: Past, Present, and Future*. Journal of Economic Literature 63 (1): 197–287. DOI: 10.1257/jel.20221319
40. Bradshaw, J., Middleton, S., Davis, A., Oldfield, N., Smith, N., Cusworth, L. and J. Williams, 2008. A minimum income standard for Britain: what people think. Joseph Rowntree Foundation. https://www.jrf.org.uk/report/minimum-income-standard-britain-what-people-think
41. Hill, K., Davis, A., Hirsch, D., and L. Marshall, 2016. Falling short: the experiences of families living below the Minimum Income Standard. Inspiring Social Change, Joseph Rountree Foundation. https://www.jrf.org.uk/report/falling-short-experiences-families-below-minimum-income-standard

9 DOCUMENT INFORMATION

Author: William James Maybury, *BSc PhD CEng*
Title: Redefining Macroeconomic Forecasting: A Framework Validation Project
Report: Axsym Engineering Publications
Series: Economy Dynamics, Concept Paper
Paper: 01
Issue: 1
Date: 30-12-2026

Document Reference

Maybury, W. J., 2025. Redefining Macroeconomic Forecasting: A Framework Validation Project. Axsym Engineering Publications, Economy Dynamics Concept Paper 01: http://www.axsym-engineering.com/publications/EconomyDynamics_ConceptPaper01.pdf

Correspondence

email: william.maybury@axsym.co.uk

Keywords

Economy Dynamics, macroeconomic forecasting, validation and verification, V&V methodology, economic modelling, empirical economics, systems engineering, causal modelling, reproducible forecasting, policy evaluation, surrogate modelling, economic behaviour, software verification, economic simulation, engineering methodology, open science, testable frameworks, economic policy tools, household finance models, price-setting mechanisms.

Funding

This work has received (a little) crowd funding. If you are interested in my work, please considered sponsoring the economy dynamics project: https://www.patreon.com/provable_economic_theory

Acknowledgements

The author extends sincere gratitude to Dick Stroud for his invaluable feedback on the readability and impact of several early drafts of this paper.