Scientific Research Publishing

Modern Economy, 2026, 17(11), 1466-1488
https://www.scirp.org/journal/me
ISSN Online: 2152-7261
ISSN Print: 2152-7245

Outlines of an E.conometric Model

Antonio Simeone[1], Marco D'Ambra[1], Paolo Savona[2]

[1]LUISS Quantum & AI Lab, LUISS Guido Carli, Rome, Italy
[2]Fondazione AVAL, Rome, Italy
Email: profsavona@gmail.com

How to cite this paper: Simeone, A., D'Ambra, M., & Savona, P. (2026). Outlines of an E.conometric Model. *Modern Economy, 17,* 1466-1488.
https://doi.org/10.4236/me.2026.1711073

Received: June 27, 2026
Accepted: November 21, 2026
Published: November 24, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

In a previous paper, the authors proposed the transition from econometrics to e.conometrics and organised the methods that make it possible into a shrub with three shared roots, namely nonlinear mathematics, higher computational power and the neurosciences, which feed two trunks whose branches, and as yet unspecified twigs, meet in a crown. La Stella (2026) analysed this framework, introducing the notions of controlled epistemological circularity, dynamic validity and algorithmic reflexivity. This paper updates the shrub, which we now call a logical shrub, with a third trunk, language, and incorporates La Stella's analysis, defining the e.conometric approach by four elements: a plurality of modelling paradigms; the systematic use of alternative, high-frequency and textual information; an adaptive and explicitly validated framework; and explainability by design. We argue that forecasting, explaining and intervening are distinct questions with distinct validation criteria, and that the three trunks play complementary roles: learning from data offers flexible tools for prediction, simulation is one of the principal routes to intervention and to the reflexive effect of forecasts, and language can capture information not yet in the numbers. The e.conometric model can therefore be expected to grow in the crown of the logical shrub, where trunks, branches and twigs meet, held together by three functions of the system: the combination of forecasts, validation and explanation. The econometric models in common use remain components of this architecture, and the contribution of the model lies in integrating them with AI methods in an auditable way.

Keywords

E.conometrics, Economic Forecasting, Causation, Policy Choices, Explainability, Validation, Auditability, Hybrid Modelling, Alternative Data, Logical Foundations, Agent-Based Models

1. Introduction

The last two decades have repeatedly exposed the limits of the forecasting models

on which economic policy relies. The global financial crisis of 2008, which led Queen Elizabeth II to ask economists why no one had seen it coming, the inflation that followed the pandemic, and the present deep weakness of sovereign debt have each appeared as a discontinuity. Two failures should be kept apart. Some of these shocks, a pandemic above all, could not have been anticipated by any economic model: they belong to the radical uncertainty discussed in Section 3. What the econometric models in common use also failed to do was to forecast the consequences of the shocks once they had occurred, as with the strength and persistence of inflation after 2021, so that economic policy has absorbed their effects, and still absorbs them, only with difficulty. It is this second failure that a better model can hope to reduce. As one of us has emphasised (Savona, 2024), such errors are not merely technical: in 2021 they delayed the anti-inflationary decisions of monetary policy, and, as Mervyn King has argued, by feeding radical uncertainty (Kay and King, 2020) they can damage the prestige on which central banks' influence over expectations rests.

In our previous paper (Simeone, D'Ambra and Savona, 2026; SDS hereinafter), we argued that the response to these failures cannot consist in adding new tools to an unchanged conception of the model. It requires a transition from econometrics, rooted in linear mathematics, in the early progress of computing and in economic theory, to e.conometrics, which draws on nonlinear mathematics and econophysics, on higher computational power and on the developments of the neurosciences.

La Stella (2026) (LS hereinafter) offered an epistemological reading of the e.conometric framework of SDS, using as a test case what SDS called the AI Econometric Accelerator (AEAI hereinafter). Without questioning the motivation of the proposal, she asked under which conditions it can produce valid knowledge, and answered with three concepts. 1) Controlled epistemological circularity describes a learning system whose revisions, including revisions of variables, model form and theoretical constraints, remain exposed to independent empirical challenge. 2) Dynamic validity requires a model to state not only its observed performance, but also its uncertainty, its domain of applicability and the signals that its reliability is deteriorating. 3) Algorithmic reflexivity names the circuit in which human behaviour, its algorithmic representation and the behaviour modified by that representation feed into one another: forecasts enter decisions, and decisions change the system that generated the data. Her conclusion is that e.conometrics may be regarded as a promising but still open stage in a research programme in the sense of Lakatos (1978), rather than the completed proclamation of a new paradigm.

Between the two papers a gap remains: SDS provides a map of the available methods, LS the conditions that any resulting model must meet, but neither sets out which components such a model must combine and how they should be arranged. This paper addresses that gap: if the conditions of validity are taken seriously as design constraints, which ingredients must an e.conometric model combine, and how must they be arranged?

The presentation of SDS also needs revisiting. Large language models, the most visible form of generative AI, have greatly extended the possibility of treating text, narratives and qualitative information as a systematic source of economic knowledge, and of deploying software "agents" that read, reason and act on it. This goes beyond data-driven learning and gives rise to a third trunk of the logical shrub, language, alongside simulation and data. The three trunks differ not in the algorithms they use but in how they come to know the economy: simulation reconstructs its mechanisms, data-driven learning extracts its regularities, and language reads what is said about it, including the narratives that may move the economy before they become numbers (Shiller, 2017) and news content that carries information on economic activity (Bybee et al., 2024). Text can be turned into series and used as data; what makes language a trunk is its object: what agents say they expect and intend, and how institutions explain their decisions, much of which no survey or statistical series records. With large language models, AI has thus re-engaged with language, which had a golden moment in the 1920s and 1930s in the philosophy of science of the Vienna Circle of Moritz Schlick. An approach is therefore e.conometric when it combines a plurality of modelling paradigms, the systematic use of alternative, high-frequency and textual information, an adaptive and explicitly validated framework, and explainability by design.

Our central argument starts from a distinction that LS draws, building on Shmueli (2010) and Guala (2006): forecasting, explaining and intervening are different questions, each with its own form of validation, namely predictive, causal and policy-oriented. Read against these questions, the three trunks are complementary rather than competing: none answers all three on its own, and the e.conometric model, if it is to exist, can be expected to grow in the crown of the logical shrub.

We are aware that contemporary econometrics already includes nonlinear, adaptive and data-rich methods, and its models remain components of the architecture proposed here. What the e.conometric model adds is not a new class of tools but their integration under explicit and auditable rules: how components enter, how their outputs are combined and constrained, and how the whole is validated. Our answer is therefore an "architecture": three trunks whose outputs meet in the crown, held together by three functions of the system, namely the combination of forecasts, validation and explanation.

The contribution of this paper is conceptual: it separates what follows from stated assumptions from what still has to be verified empirically. Behind it lies a mathematical formulation of the architecture, which will be made public once the reactions to this "manifesto" for a research programme have been gathered. Whether the integration adds value over existing hybrid systems is an empirical question, and Section 6 proposes the tests that could answer it.

2. From the AI Econometric Accelerator to the E.conometric Model

In May 2025, at a conference organised by the IIEC at LUISS Guido Carli Univer-

sity, the first results of experiments using AI techniques for forecasting were presented. The AEAI was made public in March 2026 in the paper "From Econometrics to e.conometrics" (SDS), where the dot recalls the "electronic" transition under way. SDS represented the e.conometric framework as a shrub with two trunks; the logical shrub illustrated in this paper (Figure 1) keeps its three roots, now shown feeding every trunk, and adds a third trunk, whose branches and twigs meet with those of the other two in a crown.

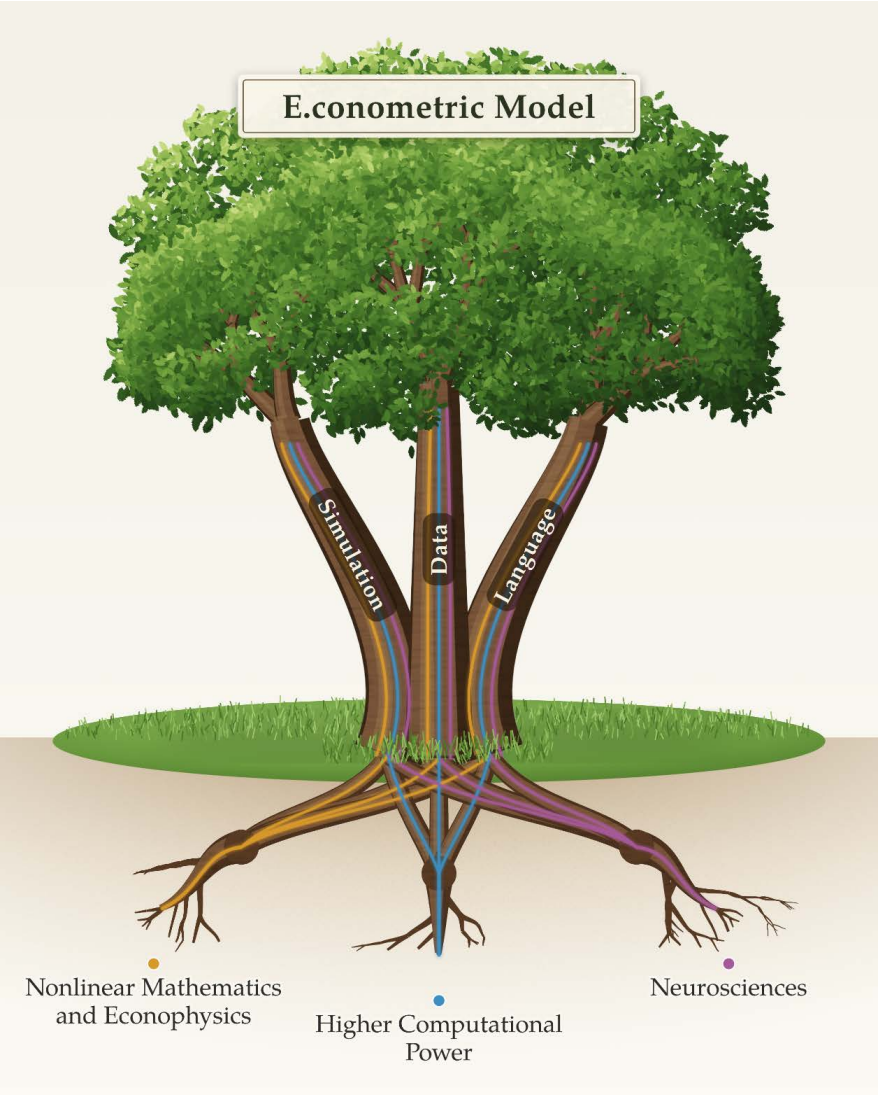


Figure 1. The logical shrub of E.conometrics: three shared roots, Nonlinear Mathematics and Econophysics, Higher Computational Power and the Neurosciences, feed three intertwined trunks, Simulation, Data and Language, whose branches and twigs meet in the crown where the E.conometric Model grows.

In SDS the established econometric model was represented as a tree: its roots are linear mathematics, economic theory and the progress of computing; its single trunk is the Classical Linear Regression Model (CLRM); and its two branches are

time-series analysis, whose central tools are vector autoregressions (VARs) and their Bayesian versions (Sims, 1980; Litterman, 1986), and structural modelling, whose most developed form is the New Keynesian dynamic stochastic general equilibrium (DSGE) model (Christiano, Eichenbaum and Evans, 2005; Smets and Wouters, 2007).

The logical shrub does not discard the old tree: it renews its roots, turns its single trunk into three and enriches its branches and twigs, creating a new crown.

The resulting map is as follows.

- **Roots**
  - Nonlinear Mathematics and Econophysics: nonlinear dynamics (chaos, self-organised criticality, collective behaviour, delay embedding) and econophysics (stylised facts, power laws, random matrices)
  - Higher Computational Power: large-scale simulation, sampling methods, differentiable programming
  - Neurosciences: behavioural (bounded rationality, prospect theory, neuroeconomics) and computational (artificial neural networks, backpropagation, attention)
- **Simulation trunk**
  - Agent-based models: macroeconomic agent-based models for forecasting, data-driven agent-based models
  - Networks and input-output: network science, Leontief propagation, network origins of aggregate fluctuations, dynamic input-output models, financial contagion
  - Learning agents: reinforcement-learning agents trained within simulators, deep hedging
- **Data trunk**
  - Statistical and machine learning: vector autoregressions and Bayesian VARs, factor models, regularised linear models, tree ensembles
  - Regimes: Markov-switching, threshold, time-varying parameter and change-point models
  - Deep learning and foundation models for time series and tables: recurrent networks, Transformers, time-series and tabular foundation models
  - Causal learning: double machine learning, causal forests, Granger causality, PCMCI, convergent cross mapping, power-law causal discovery
- **Language trunk**
  - Text as measurement: dictionaries, embeddings, language-model classification, narrative and uncertainty indices
  - Language models as readers and forecasters: conditional forecasts, chronologically consistent models
  - AI agents: tool use, multi-agent deliberation and verification
- **Crown**
  - Simulation and Data: simulation-based calibration, surrogates used to calibrate simulators to observed data, data assimilation, nowcast-initialised

simulators, graph neural networks
    - Theory and Data: physics-informed learning, DSGE priors, accounting reconciliation
    - Language and Data: text features
    - Language and Simulation: language-model agents in simulators
    - System functions: combination of predictive distributions, validation and explanation

A manifesto for a new model is credible only if it is fair to the old one. Four properties of the established family are achievements, not limitations, and any successor must retain them.

- **Identification**. Structural models and identified VARs attach economic meaning to shocks and parameters. They answer questions about mechanisms, not only about outcomes.
- **Counterfactuals**. Because their structure is explicit, they can simulate the effects of a policy that has never been observed. Provided their parameters are invariant to the policy being evaluated, this is precisely what Lucas Jr. (1976) demanded of models used for policy evaluation.
- **Accounting consistency**. Appropriately specified structural systems impose national accounting identities and budget constraints by construction. Statistical forecasts need not satisfy them, and a forecast that violates an exact identity is not merely inaccurate but internally inconsistent, and cannot serve as a coherent basis for policy. Approximate regularities, by contrast, are restrictions to be tested rather than imposed.
- **A shared language**. Their assumptions are public and their failures are diagnosable. Institutions can discuss, audit and compare them.

The limits of the established family follow from the same choices that made it tractable, and since the 2008 financial crisis they have been at the centre of a debate on how macroeconomic models should be rebuilt (Blanchard, 2018; Vines and Wills, 2018). We group them, for simplicity, into five.

- **Linearity and Gaussianity**. Many operational models are linearised around a steady state and driven by Gaussian shocks. Economic and financial data instead display heavy tails and, in financial markets, volatility clustering (Cont, 2001; Fagiolo, Napoletano and Roventini, 2008), features that matter especially in crisis episodes.
- **The representative agent (or market operator).** Representative-agent specifications reduce the variety of market operators to a single representative one, and so suppress their heterogeneity and interaction (Kirman, 1992). Existing heterogeneous-agent models address part of this problem; whether a richer representation improves forecasting or policy analysis remains an empirical question.
- **Stationarity and equilibrium**. A specification estimated under one regime may become unreliable after a break, and forecast failure then becomes systematic rather than random (Clements and Hendry, 2006). Econometric and

AI methods both face this risk; neither equilibrium modelling nor stationarity is a universal requirement of contemporary econometrics.

- **Low frequency and publication lags**. Many operational models are estimated on quarterly official data released weeks or months late, with a ragged edge of incomplete recent observations (Wallis, 1986). Mixed-frequency and state-space models address the ragged edge, but largely within official and survey statistics: the information that would reveal a turning point may arrive too late to be used.
- **A narrow information set.** Unstructured, high-frequency and alternative information, from transactions to text, has no natural place in a system of structural equations, and incorporating it requires measurement and alignment rules that the established family was not designed to provide.

These issues have generated established research programmes, including heterogeneous-agent models (Kaplan, Moll and Violante, 2018), time-varying parameter VARs (Primiceri, 2005), nonlinear solution methods and factor-model nowcasting (Giannone, Reichlin and Small, 2008). Each addresses one or two limits at a time, often at a cost in tractability, the very constraint the established family was built around. What we propose is not another extension but an explicit architecture in which components addressing different limits are connected and their joint use is audited.

3. Widening the Logical Scope of the E.conometric Model: Forecasting, Explaining, Intervening

LS has widened the scope of this research by raising the point that the use of forecasts cannot be separated from their validation, nor from an assessment of how they are used in the light of what they are meant to achieve.

One limit cannot be overcome by any model, old or new. Knight (1921) distinguished risk, which admits a probability distribution, from uncertainty, which does not. As LS observes, novel events and regime changes may not belong to a sufficiently homogeneous class of past cases, and the task is not to absorb uncertainty into calculation, but to build models that can tell what they can represent from what exceeds them.

The debate between structural and data-driven modelling is often framed as a contest over accuracy. It is better framed, as LS suggests, as a difference in the questions being asked. Shmueli (2010) argued that explanation and prediction are often conflated although they pursue different goals and require different criteria; Breiman (2001) described a closely related divide between two cultures of statistical modelling, one assuming a stochastic data model and one relying on algorithmic models judged mainly by predictive accuracy. In economics, Mullainathan and Spiess (2017) drew the same line for machine learning, which is built to predict rather than to recover the parameters that policy analysis requires (Athey and Imbens, 2019). Guala (2006) adds a third question that economics cannot avoid, because its models are used to guide action: what would happen if we intervened?

The three questions are not ranked. A good forecast can be useful without an explanation, and a credible causal estimate can be insufficient for policy. What matters is that each requires its own form of validation; Table 1 builds on what LS calls a validation chain.

The third row of Table 1 is where the Lucas Jr. (1976) critique arose, and where data-driven models reopen it. He argued that relationships estimated under one policy regime cannot in general be used to evaluate alternative policies, because the decision rules they embody change when policy changes. Structural models answered by deriving behaviour from deep parameters, whose invariance to policy is itself an assumption. Data-driven models, which learn reduced-form relationships by construction, reopen the critique, and in a stronger form when they are continuously retrained on behaviour that their own forecasts have influenced.

Table 1. Three questions and their forms of validation.

Question	What it asks	Validation	Typical failure
Forecasting	What will happen, or what is happening now?	Predictive: out-of-sample accuracy, calibrated uncertainty, stability across regimes	Accurate on the past, fragile after a break
Explaining	Why does it happen, and which drivers move it?	Causal: identification assumptions, robustness to confounding, consistency with theory	Spurious association read as mechanism
Intervening	What would happen if we acted?	Policy: counterfactual simulation, behavioural responses, distribution of effects, asymmetric losses	Relationships estimated without the policy break down once it is applied

LS names the broader mechanism of algorithmic reflexivity: human behaviour, its algorithmic representation and the behaviour modified by that representation enter the same circuit. A forecast is published, it shapes expectations and decisions, and those decisions generate the data on which the next forecast will be trained. The question is older than the Lucas critique. Morgenstern (1928) argued that economic forecasting is impossible in principle, partly because agents react to published forecasts; Grunberg and Modigliani (1954) and Simon (1954) showed that, if agents' reaction to a published forecast is known and continuous, there exists a forecast that remains correct once agents have reacted to it; and Goodhart (1975) observed that a regularity used for control tends to collapse. Algorithmic reflexivity extends the question from a single published forecast to models that are continuously retrained on the behaviour they influence. The machine-learning literature studies a closely related problem under the name of performative prediction (Perdomo et al., 2020). Drawing on Mervyn King, one of us has argued that forecasting errors feed radical uncertainty (Kay and King, 2020) and damage the prestige of central banks, the main basis of their influence on market expectations (Savona, 2024). Once markets perceive such errors as systematic, the credibility of monetary-policy choices is itself at stake.

The consequence for model design is direct. A model that learns only from historical data absorbs the effects of its past forecasts only after they have occurred, mixed with everything else, and cannot represent the effect of the forecast it is about to publish. To represent this effect, the forecast must enter agents' information and decision rules, so that its feedback effects are represented explicitly. Agent-based simulation does this most directly; structural models with endogenous expectations and causal models of forecast deployment can also do it under stated assumptions. An observational predictive learner alone does not identify the counterfactual effect of publishing its next forecast.

The three questions do not map one-to-one onto the three trunks. Data-driven learning supplies flexible forecasting tools, but its superiority is task-dependent. Simulation represents interventions under explicit behavioural assumptions and can also forecast when adequately calibrated. Explanation requires justified causal reasoning across the architecture. Language models can propose hypotheses and measurements but cannot validate a causal claim merely by generating a plausible account.

Language can contribute to all three questions, each in a different way: it can detect qualitative shocks before they appear in the series, which serves forecasting; it can track how expectations are formed and communicated, which matters for intervention; and it is the medium through which explanations, produced by causal and simulation tools, reach decision-makers. No trunk therefore answers all three questions on its own, and their limits differ: what one leaves open, another can partly address. This is not an argument for eclecticism but a structural one: a model that must forecast, explain and support intervention can be expected to draw on all three ways of knowing the economy, with components connected so that each can compensate for the blind spots of the others. The requirements in the next section make this argument precise.

4. Definition and Requirements of the E.conometric Model

Building on the definition of SDS, we define an approach as e.conometric when it combines four elements. The first three answer limits of the established model discussed in Section 2; the fourth preserves one of its achievements, the shared language discussed there, and extends it:

1) a plurality of modelling paradigms, drawn from simulation, learning from data and language, none of which needs to start from a pre-specified system of structural equations;

2) the systematic use of alternative, high-frequency and textual information alongside official statistics;

3) an adaptive and explicitly validated framework, in which models, data sources and validation procedures are continuously compared and revised, and every output is reported with its uncertainty and the domain in which it holds; and

4) explainability by design, so that every output is accompanied by an attribution of its drivers under a declared method, traceable to the data and model ver-

sions that produced it, and stated in terms that a decision-maker can audit.

The first generalises SDS's first criterion, which required nonlinear, nonparametric or data-driven techniques, into a plurality of paradigms drawn from all three trunks, whose value as a robustness mechanism LS emphasises. The second makes explicit the textual information that SDS counted among alternative data. The third and fourth help separate a research programme from a toolkit, the economist's "box of tools" in Joan Robinson's phrase (Robinson, 1933). The third, which we take from LS, keeps the model exposed to independent challenge; the fourth, which this paper adds to the definition, makes its output usable by an institution, which must be able to say why a forecast moved, as when a nowcast is revised after new data are released, and on what it rests.

The plurality required by the first element is a plurality of trunks, not merely of techniques within a single trunk. We distinguish the e.conometric approach, which is defined by the four elements above, from implementations of the proposed architecture. An implementation integrating Simulation and Data under explicit combination, validation and explanation rules is a legitimate partial implementation, but not the complete architecture proposed here. The complete e.conometric architecture requires all three trunks: Simulation, Data and Language. Simulation and Data between them already serve every requirement set out below, at least in part. Neither, however, fully captures what agents and institutions say about the economy: expectations, intentions and narratives that language can read before they appear in numerical series. The Language trunk therefore completes the proposed architecture by providing a distinct information channel, rather than merely another input to the Data trunk.

E.conometrics is therefore not a synonym for machine learning in econometrics, for agent-based modelling or for nowcasting: each is a possible ingredient, and the distinctive claim lies in their possible integration. Nor does it give up what the structural tradition achieved: identification and counterfactual analysis are retained, both in the structural components that remain part of the architecture and through causal learning, the simulation of mechanisms and theory used as a constraint. Nor does it discard economic theory. Theory moves from the foundation of the model to one of its disciplines, acting as a constraint, a prior or a benchmark against which what the model learns is checked. A typical example of theory that should be tested rather than assumed is the idea that inflation always has monetary origins, rooted in economic doctrine as well as in popular culture, from which distorted behaviours arise, putting burdens on monetary-policy choices; fiscal policy, laden with ideological assumptions, faces the same risk.

The definition says what the e.conometric model is; the requirements state what it must do. The requirements come from two sources: the achievements and limits of the established model, and the conditions of validity proposed by LS. Each is phrased as a property of the architecture as a whole, not of a single technique or component, so that different techniques can be judged against it. **Table 2** lists them.

Table 2. The decalogue of requirements of the e.conometric architecture.

#	Requirement	What it demands	Source	Main question served
R1	Timeliness	Estimate the present and near future from mixed-frequency, asynchronous data, resolving the ragged edge	Publication lags, low frequency; narrow information set	Forecasting
R2	Nonlinearity and regime awareness	Represent amplification, thresholds and heavy tails; detect regime changes and adapt to them, with detection evaluated under declared conditions	Linearity, stationarity	Forecasting
R3	Heterogeneity and interaction	Represent heterogeneous agents and the networks through which shocks propagate	Representative agent (or market operator)	Intervening
R4	Theoretical coherence	Enforce accounting identities where they hold exactly; use empirical regularities as soft penalties, priors or testable benchmarks, not as exact restrictions	Preserve accounting consistency	All
R5	Causal grounding and identification	Distinguish drivers from correlates, and state the identifying assumptions on which every causal relationship rests	Forecasting vs explaining; preserve identification	Explaining
R6	Interpretability	Decompose each output into drivers that a decision-maker can audit	Preserve a shared language	Explaining
R7	Calibrated uncertainty	Produce distributions rather than points, with scores, coverage and tails assessed out of sample, stating the conditions of any calibration guarantee	Gaussian shocks; point forecasts	All
R8	Counterfactual and reflexive capacity	Simulate interventions, including agents' reactions to policy and to the forecast itself, with behavioural assumptions stated and checked against identified evidence where it exists	Preserve counterfactuals; Lucas critique, algorithmic reflexivity	Intervening
R9	Controlled revision	Revise parameters, variables, form and constraints, versioning every revision and exposing it to out-of-sample challenge under strict temporal separation: real-time data releases and a documented training chronology for pre-trained models	Controlled epistemological circularity; temporal separation	All
R10	Dynamic validity and abstention	State performance, uncertainty, domain of applicability and signals of deterioration; degrade gracefully, and abstain as a last resort, when outside the validated domain	Dynamic validity; Knightian uncertainty	All

Three remarks on the list.

First, the requirements are unevenly served: some are met natively by one trunk and left open by the others, and five of them, R5, R6, R7, R9 and R10, are met fully by no trunk (Table 4). R5 and R7 require validation beyond any model class, causal validation for the first and calibration assessment for the second; combination contributes to both but completes neither by itself. R6 requires an attribution protocol and a faithful account of the evidence, which no trunk supplies alone. R9 and R10 concern the architecture as a whole rather than any of its components.

Second, R9 and R10 turn validation from a phase that follows modelling into a function of the model itself. This is LS's central point, and it has a design consequence: monitoring, comparison between components and the decision to abstain must be built into the system, not added afterwards.

Third, the requirements are not independent. Plurality of components serves R7 and R10 at the same time: as LS observes, when representations built on different assumptions converge, confidence may be strengthened, and when they diverge, the divergence may itself be information about structural uncertainty. A single model can also represent parameter or structural uncertainty; plurality adds diagnostics, it does not replace them. The signal is informative only to the extent that the components differ in assumptions and data: agreement can also reflect shared errors, and its reliability must be evaluated rather than presumed.

Each requirement must also be assessable. Table 3 states, for each, the operational evidence or mathematical condition that would support the claim that it is met; no cell of Table 4 should be read as unconditional satisfaction of these tests. The role of every language component must be recorded rather than inferred from its label, since language can act as a measurement channel, a forecasting component or a behavioural module inside a simulator. Targets must likewise be declared exactly, including their transformation, horizon and measurement convention: differently defined targets are not comparable.

Table 3. Operational tests of the requirements.

#	Requirement	Operational evidence or mathematical condition
R1	Timeliness	Timestamped releases and reproducible forecasts from the archive available at each origin, with accuracy reported by information date relative to official releases.
R2	Nonlinearity and regime awareness	Pre-specified diagnostics and performance reported across declared regimes; nonlinearity alone is insufficient.
R3	Heterogeneity and interaction	Recorded agent states, interaction structure and sensitivity to heterogeneity assumptions.
R4	Theoretical coherence	Explicit feasible sets and constraint residuals; orthogonal projection is guaranteed not to increase the total squared error only when the target satisfies the restrictions.

Continued

R5	Causal grounding and identification	A causal estimand, identification assumptions and a defensible estimation design; predictive accuracy is insufficient.
R6	Interpretability	An auditable chain from inputs and model versions to output, separating associations from causal mechanisms.
R7	Calibrated uncertainty	Proper scores, interval width and realised coverage by horizon and regime; mixtures alone do not guarantee calibration.
R8	Counterfactual and reflexive capacity	A stated intervention mechanism, behavioural assumptions and comparison with identified policy evidence.
R9	Controlled revision	Frozen historical forecasts, release records and chronological validation of each change.
R10	Dynamic validity and abstention	Pre-specified monitoring, fallback and abstention rules, evaluated together with their frequency and decision loss.

The requirements are thus obligations of design and evaluation: some can be supported by conditional mathematical results, others require empirical evidence.

We can now return to the claim that no trunk can give a satisfactory answer on its own, but only jointly with the others. Table 4 assesses each trunk on its own, without hybrids. A black cell marks a requirement that the trunk typically serves natively; a grey cell, one it meets only in specific settings or with additional tools; a white cell, one it leaves open.

Table 4. The decalogue requirements served by each trunk on its own.

Requirement	Simulation	Data	Language
R1 Timeliness			
R2 Nonlinearity and regime awareness			
R3 Heterogeneity and interaction			
R4 Theoretical coherence			
R5 Causal grounding and identification			
R6 Interpretability			
R7 Calibrated uncertainty			
R8 Counterfactual and reflexive capacity			
R9 Controlled revision			
R10 Dynamic validity and abstention			

Simulation most directly serves heterogeneity and interaction (R3) and counterfactual and reflexive capacity (R8), since an intervention or a published forecast can be introduced into the agents' information set and its effects simulated; it can impose accounting consistency by construction (R4) and generates nonlinear dynamics and regime changes endogenously (R2). It is calibrated, however, on national accounts that arrive late (R1), and its computational cost makes controlled revision (R9) so expensive that in practice it is left open. Learning from data is the most direct answer to timeliness (R1) and nonlinearity (R2), and has the most established practice of out-of-sample evaluation, which gives it partial tools for R9 and R10; causal learning and models with density outputs contribute to R5 and R7, but predictive models learned from history cannot represent new interventions or their own reflexive effect (R8). Language reads shocks visible in news, decisions and statements before they appear in the series, which serves R1 and the qualitative side of R2, and is the medium through which explanations reach decision-makers (R6); but its outputs are not calibrated probabilities unless they are made so (R7), it is exposed to look-ahead bias (R9), and it can propose causal stories without being able to validate them (R5).

The strengths of simulation and data are largely complementary, and hybrids between trunks can fill several of the remaining gaps. The shading of the table describes typical capabilities rather than unconditional guarantees. None of the trunks or their combination automatically establishes identification, calibration or dynamic validity. R5 and R7 need distinct causal and probabilistic validation, R6 needs an attribution protocol, and R9 and R10 govern the complete deployed process.

These gaps are what the three functions of the system address. Combination merges the predictive distributions of comparable components, with weights based on their past performance, and treats disagreement among them as information about structural uncertainty rather than as noise to be averaged away. Several combination rules are compatible with the architecture, and the choice among them is itself an empirical question, to be settled on a chronologically separate validation sample. One candidate, given here for concreteness, is an optimal prediction pool (Geweke and Amisano, 2011): at each forecast origin, the weights are those that maximise the average log predictive score of the combined distribution, that is, the logarithm of the density it assigned to the outcomes actually observed, over a rolling window of recent origins, for instance the last eight quarters for a quarterly target, using only outcomes already released. A longer window yields more stable weights, a shorter one adapts faster after a regime change. Alternatives include equal weights, as a benchmark, and weights updated exponentially in the components' realised losses, which carry guarantees relative to the best single component (Cesa-Bianchi and Lugosi, 2006). Validation is built into the model rather than added afterwards: calibration is monitored, pre-specified signals trigger revision, and the model degrades gracefully and, as a last resort, abstains; causal claims are reported with their identifying assumptions, and simulated re-

sponses to policy are checked against episodes in which the policy did occur. Because validation carried out only by those who build the model may fall short of the independent challenge that LS requires, it is complemented by an independent assessment of the deployed system, repeated at fixed intervals and whenever a revision is triggered. Explanation, finally, attributes each output to its drivers under a declared method, separates associations from causal mechanisms and, to ensure auditability, accompanies every output with an audit record, covering data sources and release dates, model versions, weights and diagnostics, that allows an external party to reproduce it.

The architecture can be compared with two established bodies of practice. Hybrid forecasting combines models of different kinds: statistical and machine-learning components, as in hybrids of ARIMA models and neural networks (Zhang, 2003) or of exponential smoothing and recurrent networks, the method that won the M4 competition (Smyl, 2020); theory and data, as when a DSGE model supplies the prior for a VAR (Del Negro and Schorfheide, 2004); and, more broadly, the combination of forecasts from different models (Wang et al., 2023). Model governance prescribes how models are documented, validated and challenged: supervisory guidance on model risk management requires an evaluation of conceptual soundness, ongoing monitoring, outcomes analysis and effective challenge by parties independent of model development (Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency, 2011); documentation standards and internal audit frameworks pursue similar aims for machine-learning systems (Mitchell et al., 2019; Raji et al., 2020); and the review of forecasting at the Bank of England has called for forecast errors and their sources to be analysed systematically and for models to be revised when structural change is suspected (Bernanke, 2024). E.conometrics draws on both traditions. Hybrid systems can already integrate multiple model families, textual information and validation procedures; neither the number of components nor the use of text alone establishes novelty. The distinctive proposal here is to organise Simulation, Data and Language as three complementary ways of knowing the economy within an explicitly specified and auditable architecture. Its combination rules, explanation protocols and validation obligations are defined as system functions, with controlled revision under strict temporal separation (R9) and dynamic validity with abstention (R10) applying across components and revisions. This internal integration does not replace model-risk governance: the independent assessment described above remains necessary. Whether these design obligations yield measurable improvements over strong existing hybrid systems is an empirical question that Section 6 is designed to answer.

5. From Traditional to Alternative Data

The e.conometric model does not replace the information set of the established model; it extends it. This section describes what the extension adds, what it costs, and why managing those costs belongs to the model rather than to a preliminary

cleaning stage.

5.1. The Traditional Information Set

Official statistics, such as national accounts, industrial production, prices, labour-market surveys and business and consumer confidence, have four virtues: long histories that span several business cycles, stable and largely internationally harmonised definitions, institutional quality control, and a direct correspondence with the aggregates that policy targets. Their weaknesses follow from the same design. They are mostly monthly or quarterly, they are released with lags of weeks to months, they are revised after publication (Croushore, 2011), and they measure only what statistical offices decided in advance to count.

5.2. What Alternative Data Add

Alternative data are produced for purposes other than statistics, often as a by-product of economic activity. The main families now used in macroeconomic monitoring are:

- **Payments and transactions:** card and other electronic payment flows, which improve nowcasts of GDP and its main components (Aprigliano, Ardizzi and Monteforte, 2019) and were used by researchers, during the pandemic, to build public real-time trackers (Chetty et al., 2024).
- **Energy and physical activity**: electricity and gas consumption, which can track production at daily frequency.
- **Transport and mobility**: location-based measures of commuting, retail and travel activity, and freight and port data.
- **Remote sensing**: satellite imagery, notably night-time lights, which can complement official measures of growth where national accounts are weak (Henderson, Storeygard and Weil, 2012).
- **Weather and climate**: observed and forecast temperature, precipitation, wind and solar radiation, and reanalyses that reconstruct them over decades, used to forecast energy demand and renewable supply (De Felice, Alessandri and Ruti, 2013; Antonini et al., 2024).
- **Web activity:** search volumes (Choi and Varian, 2012) and online prices (Cavallo and Rigobon, 2016).
- **Text:** news, filings, central-bank communication and social media. Text is also data, but it is the object of the third trunk.

High-frequency composites such as the Weekly Economic Index (Lewis et al., 2022) registered the collapse of activity in March 2020 within days, weeks before the first official quarterly estimate was released.

5.3. The Costs

The gain in timeliness is paid for in five "currencies".

- **Frequency mismatch**. Daily and weekly series must inform a quarterly target, and no single aggregation rule is neutral.

- **The ragged edge**. At any date, each series ends at a different point (Wallis, 1986). The available information set is unbalanced at its most recent end, which is the end that matters.
- **Short histories**. Many alternative sources begin only in the 2000s or 2010s and therefore span few business cycles. Methods that need long samples, flexible nonlinear learners in particular, are prone to overfitting in short samples.
- **Representativeness and drift**. Data come from a provider's customers, not from a designed sample. Coverage changes with market shares, technology and behaviour, so the measuring instrument itself can shift.
- **Access and privacy**. Many sources are proprietary, licensed and subject to data-protection rules, which limits replication and constrains how granular inputs can be.

Each of these costs has established mitigations, none of which removes it entirely. Alignment, imputation and quality control are not neutral pre-processing: they determine what the model sees, and they can introduce artefacts as easily as they remove noise. They must therefore be versioned, validated out of sample and revised like any other component, which is requirement R9 applied to data.

6. This Manifesto as an Open Research Agenda

In Lakatos's (1978) terms, a research programme is progressive when it predicts novel facts, some of which are then corroborated. Table 5 therefore states claims that can fail empirically. Key tests concern the incremental value of hybrids, information chronology and the decision usefulness of monitoring. Two of these tests are particularly critical, because their outcome could put the programme in difficulty: hybrids are expected to outperform the best single trunk where Table 4 indicates complementary gaps, and pre-specified signals are expected to anticipate forecast failure more often than a benchmark rule; if they do not, the central claims of this paper are weakened. The preliminary evidence reported in SDS must be replicated with real-time archives and admissible foundation-model training chronology, as requirement R9 demands.

Table 5. An agenda of empirically assessable steps.

Step	Question	Evidence of progress
Real-time benchmarks	Do e.conometric models beat established nowcasts when evaluated on the data actually available at each date?	Real-time comparisons against official and survey nowcasts, reported separately for calm, crisis and recovery regimes
Leakage-free foundation models	How much of the reported gain of foundation models and language models survives once temporal contamination and revision bias are removed?	Models pre-trained on synthetic or time-stamped corpora, evaluated against their unconstrained counterparts
Hybrid versus single trunk	Does a hybrid outperform each of its components where Table 4 indicates complementary gaps?	Nowcast-initialised simulators against both the nowcast and the standard simulator; identity-constrained learners against unconstrained ones at turning points

Continued

Reflexivity	How large are the second-order effects of publishing a forecast, in simulation and in data?	Simulations in which forecasts enter the information set of learning or language-model agents, compared with outcomes when they do not; where available, natural experiments around changes in forecast publication
Dynamic validity	Do pre-specified signals anticipate forecast failure, and is abstention valuable?	Coverage of predictive intervals, calibration over time, and the loss avoided by widening or withholding forecasts when signals deteriorate
Explanation fidelity	Do explanations describe the model's actual response to changes in its inputs, and are they consistent across methods and with causal evidence?	Response of the model's output to controlled changes in the attributed inputs, compared with the attributions; agreement across methods and model components, consistency with causal discovery and simulated mechanisms, and stability between adjacent forecasts, as supporting diagnostics
Causal convergence	Do causal conclusions agree across trunks?	Agreement between causal discovery on data, mechanisms simulated in agent-based models and drivers identified in text, read as a diagnostic rather than as proof of causation
Policy validation	Do simulated responses to policy match the evidence from past policy episodes?	Comparisons with responses estimated by local projections and structural decompositions, and models trained before a policy change evaluated after it
Open infrastructure	Can results be reproduced across institutions?	Open simulators and benchmarks, shared real-time data archives, and privacy-preserving access to alternative data

Among the further questions LS raises, two bear directly on this programme: the first concerns the nature of economic decisions in the light of the neurosciences and of emotion, which the behavioural root of the logical shrub only begins to address; the second concerns generative intelligence: a system that produces scenarios and hypotheses expands the space of what can be considered, but its output becomes knowledge only when it is brought back to control, comparison and validation.

Protocol for Testing the Architecture

Theoretical guarantees and empirical performance are separate claims. Evaluation must archive origin, target, horizon, release, model version, raw component outputs, integration weights and all diagnostics. Preprocessing is fitted and hyperparameters are tuned inside each chronological training period. Benchmarks include naive and autoregressive forecasts, competitive mixed-frequency or bridge models, factor models and regularised or nonlinear learners, compared under matched information budgets. Institutional forecasts need matching publication dates and

target conventions; their method cannot be inferred from their institutional name.

For point forecasts of a scalar target, the mean absolute error (MAE) and the root mean squared error (RMSE) are reported by horizon, together with individual errors when the evaluation sample is small, and any alternative error measure is labelled explicitly. Paired loss differences need dependence-aware uncertainty analysis, adequate samples and disclosure of benchmark searches and multiple comparisons. Selecting the best of a finite library of models on the same validation sample creates optimism, which the evaluation must account for. A small number of evaluation outcomes cannot support a strong general superiority claim.

Ablations compare matched versions with and without language, alternative numerical data, simulation, reconciliation and adaptive weighting. This separates information gains from architectural gains. Probabilistic comparisons use an appropriate proper score as well as coverage and width; a discrete simulation or constrained distribution may require scores valid on its support rather than a density log score. Explainability tests assess faithfulness and stability. Faithfulness is a property of the explanation with respect to the model: the drivers to which an output is attributed must describe how the model's output actually responds when those inputs are changed in a controlled way, for instance when each in turn is replaced by its baseline value while the others are held fixed, and the size and sign of the resulting change are compared with those of the attribution. Agreement between attribution methods and stability between adjacent forecasts are useful diagnostics but do not establish faithfulness, and neither establishes a causal effect in the economy. Decision usefulness is assessed against a declared comparison policy: the decisions that the same decision rule, fixed in advance, would have produced from a benchmark forecast, such as the institution's existing one, with losses scored under a loss function also fixed in advance. A lower realised loss can be attributed to the architecture only under stated assumptions: that the decision rule is the same in both cases, and that the outcome against which both decisions are scored would not itself have changed had the other decision been taken. The second assumption is plausible for decision-makers too small to move the variables they forecast; for a central bank or a government, whose decisions shape the outcome, the comparison is counterfactual and must rely on the structural or simulated representations of policy discussed in Section 3.

7. Conclusion

This "Manifesto" extends the e.conometric approach of SDS into an architecture to forecast, explain and support decisions, addressing the three questions distinguished by LS, each with its own form of validation, to which Simulation, Data and Language contribute different information and modelling capabilities. Their complementarity is a design hypothesis rather than a general superiority theorem, and structural econometric models remain eligible components. Intervention and reflexivity require explicit behavioural mechanisms and identification, whether implemented through simulation or another structural representation. To this

end the paper adds language as a third trunk, defines the e.conometric approach by four elements and the model by ten assessable requirements of its architecture, and maps the trunks of the logical shrub onto them (Table 4). The requirements that no trunk meets on its own are assigned to hybrids and to the system functions of combination, validation and explanation. Section 6 sets out a research agenda that puts each of these claims to empirical test.

The model grows in the crown through auditable integration, economic restrictions and permanent validation and explanation. Predictive distributions, disagreement diagnostics and fallback actions must themselves be evaluated.

LS's controlled circularity, dynamic validity and reflexivity become explicit design obligations. The architecture makes them examinable, but whether they are satisfied must be shown empirically by implementations. Remaining work concerns identification, uncertainty under changing distributions, empirical gains, source chronology and whether the integration adds value over existing hybrid systems.

The aim is not a model that makes judgement unnecessary. It is a model that can give decision-makers earlier information, state how far that information can be trusted, and warn them when the conditions under which it was learned no longer hold. In an economy that, as LS puts it, changes while we attempt to understand it, this is what responsibility in the use of forecasts requires.

This is quite different from the widespread claim that AI amounts to entrusting machines with deciding in place of human beings, which can happen only if humans abdicate the role they must in any case play. Herein lies our objection to an ethics of AI built on Luddite foundations.

Author Contributions

Antonio Simeone, Marco D'Ambra and Paolo Savona jointly contributed to the conceptualization, development and preparation of this paper. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Antonini, E. G. A., Di Bella, A., Savelli, I., Drouet, L., & Tavoni, M. (2024). Weather- and Climate-Driven Power Supply and Demand Time Series for Power and Energy System Analyses. *Scientific Data, 11,* Article 1324. https://doi.org/10.1038/s41597-024-04129-8

Aprigliano, V., Ardizzi, G., & Monteforte, L. (2019). Using Payment System Data to Forecast Economic Activity. *International Journal of Central Banking, 15,* 55-80.

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know about. *Annual Review of Economics, 11,* 685-725. https://doi.org/10.1146/annurev-economics-080217-053433

Bernanke, B. S. (2024). *Forecasting for Monetary Policy Making and Communication at the Bank of England: A Review.* Bank of England.

https://www.bankofengland.co.uk/independent-evaluation-office/forecasting-for-monetary-policy-making-and-communication-at-the-bank-of-england-a-review

Blanchard, O. (2018). On the Future of Macroeconomic Models. *Oxford Review of Economic Policy, 34,* 43-54. https://doi.org/10.1093/oxrep/grx045

Board of Governors of the Federal Reserve System, & Office of the Comptroller of the Currency (2011). *Supervisory Guidance on Model Risk Management* (SR Letter 11-7). https://www.federalreserve.gov/boarddocs/srletters/2011/sr1107.pdf

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author). *Statistical Science, 16,* 199-231. https://doi.org/10.1214/ss/1009213726

Bybee, L., Kelly, B., Manela, A., & Xiu, D. (2024). Business News and Business Cycles. *The Journal of Finance, 79,* 3105-3147. https://doi.org/10.1111/jofi.13377

Cavallo, A., & Rigobon, R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives, 30,* 151-178. https://doi.org/10.1257/jep.30.2.151

Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press. https://doi.org/10.1017/cbo9780511546921

Chetty, R., Friedman, J. N., Stepner, M., Alatas, H., Baker, C., Barnhard, H. et al. (2024). The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data. *The Quarterly Journal of Economics, 139,* 829-889. https://doi.org/10.1093/qje/qjad048

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record, 88,* 2-9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Christiano, L., Eichenbaum, M., & Evans, C. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy, 113,* 1-45. https://doi.org/10.1086/426038

Clements, M. P., & Hendry, D. F. (2006). Chapter 12 Forecasting with Breaks. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (pp. 605-657). Elsevier. https://doi.org/10.1016/s1574-0706(05)01012-8

Cont, R. (2001). Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance, 1,* 223-236. https://doi.org/10.1080/713665670

Croushore, D. (2011). Frontiers of Real-Time Data Analysis. *Journal of Economic Literature, 49,* 72-100. https://doi.org/10.1257/jel.49.1.72

De Felice, M., Alessandri, A., & Ruti, P. M. (2013). Electricity Demand Forecasting over Italy: Potential Benefits Using Numerical Weather Prediction Models. *Electric Power Systems Research, 104,* 71-79. https://doi.org/10.1016/j.epsr.2013.06.004

Del Negro, M., & Schorfheide, F. (2004). Priors from General Equilibrium Models for VARs. *International Economic Review, 45,* 643-673. https://doi.org/10.1111/j.1468-2354.2004.00139.x

Fagiolo, G., Napoletano, M., & Roventini, A. (2008). Are Output Growth-Rate Distributions Fat-Tailed? Some Evidence from OECD Countries. *Journal of Applied Econometrics, 23,* 639-669. https://doi.org/10.1002/jae.1003

Geweke, J., & Amisano, G. (2011). Optimal Prediction Pools. *Journal of Econometrics, 164,* 130-141. https://doi.org/10.1016/j.jeconom.2011.02.017

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The Real-Time Informational Content of Macroeconomic Data. *Journal of Monetary Economics, 55,* 665-676. https://doi.org/10.1016/j.jmoneco.2008.05.010

Goodhart, C. A. E. (1975). Problems of Monetary Management: The U.K. Experience. In *Papers in Monetary Economics* (Vol. I). Reserve Bank of Australia.

Grunberg, E., & Modigliani, F. (1954). The Predictability of Social Events. *Journal of Political Economy, 62,* 465-478. https://doi.org/10.1086/257604

Guala, F. (2006). *Filosofia dell'economia. Modelli, causalità, previsione*. Il Mulino.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review, 102,* 994-1028. https://doi.org/10.1257/aer.102.2.994

Kaplan, G., Moll, B., & Violante, G. L. (2018). Monetary Policy According to Hank. *American Economic Review, 108,* 697-743. https://doi.org/10.1257/aer.20160042

Kay, J., & King, M. (2020). *Radical Uncertainty: Decision-Making for an Un-Knowable Future*. The Bridge Street Press.

Kirman, A. P. (1992). Whom or What Does the Representative Individual Represent? *Journal of Economic Perspectives, 6,* 117-136. https://doi.org/10.1257/jep.6.2.117

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Houghton Mifflin.

La Stella, L. (2026). From Econometrics to E.conometrics: Models, Forecasting, and Conditions of Validity. *Modern Economy, 17,* 1238-1253. https://doi.org/10.4236/me.2026.179059

Lakatos, I. (1978). *The Methodology of Scientific Research Programmes: Philosophical Papers Volume 1* (J. Worrall, & G. Currie, Eds.). Cambridge University Press.

Lewis, D. J., Mertens, K., Stock, J. H., & Trivedi, M. (2022). Measuring Real Activity Using a Weekly Economic Index. *Journal of Applied Econometrics, 37,* 667-687. https://doi.org/10.1002/jae.2873

Litterman, R. B. (1986). Forecasting with Bayesian Vector Autoregressions—Five Years of Experience. *Journal of Business & Economic Statistics, 4,* 25-38. https://doi.org/10.1080/07350015.1986.10509491

Lucas Jr., R. E. (1976). Econometric Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy, 1,* 19-46. https://doi.org/10.1016/s0167-2231(76)80003-6

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B. et al. (2019). Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220-229). ACM. https://doi.org/10.1145/3287560.3287596

Morgenstern, O. (1928). *Wirtschaftsprognose: Eine Untersuchung ihrer Voraussetzungen und Möglichkeiten*. Julius Springer.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives, 31,* 87-106. https://doi.org/10.1257/jep.31.2.87

Perdomo, J., Zrnic, T., Mendler-Dünner, C., & Hardt, M. (2020). Performative Prediction. *Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research, 119,* 7599-7609.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies, 72,* 821-852. https://doi.org/10.1111/j.1467-937x.2005.00353.x

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B. et al. (2020). Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability,*

and Transparency (pp. 33-44). ACM. https://doi.org/10.1145/3351095.3372873

Robinson, J. (1933). *The Economics of Imperfect Competition*. Macmillan.

Savona, P. (2024). *"Economic Forecasts, Errors and Financial Instability" Speech at the Trento Economics Festival, 24 May 2024*. CONSOB. https://www.consob.it/web/consob-and-its-activities/w/speech-by-president-paolo-savona-24-may-2024

Shiller, R. J. (2017). Narrative Economics. *American Economic Review, 107,* 967-1004. https://doi.org/10.1257/aer.107.4.967

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science, 25,* 289-310. https://doi.org/10.1214/10-sts330

Simeone, A., D'Ambra, M., & Savona, P. (2026). From Econometrics to E.conometrics. *Modern Economy, 17,* 464-480. https://doi.org/10.4236/me.2026.173025

Simon, H. A. (1954). Bandwagon and Underdog Effects and the Possibility of Election Predictions. *Public Opinion Quarterly, 18,* 245-253. https://doi.org/10.1086/266513

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica, 48,* 1-48. https://doi.org/10.2307/1912017

Smets, F., & Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review, 97,* 586-606. https://doi.org/10.1257/aer.97.3.586

Smyl, S. (2020). A Hybrid Method of Exponential Smoothing and Recurrent Neural Networks for Time Series Forecasting. *International Journal of Forecasting, 36,* 75-85. https://doi.org/10.1016/j.ijforecast.2019.03.017

Vines, D., & Wills, S. (2018). The Rebuilding Macroeconomic Theory Project: An Analytical Assessment. *Oxford Review of Economic Policy, 34,* 1-42. https://doi.org/10.1093/oxrep/grx062

Wallis, K. F. (1986). Forecasting with an Econometric Model: The 'Ragged Edge' Problem. *Journal of Forecasting, 5,* 1-13. https://doi.org/10.1002/for.3980050102

Wang, X., Hyndman, R. J., Li, F., & Kang, Y. (2023). Forecast Combinations: An over 50-Year Review. *International Journal of Forecasting, 39,* 1518-1547. https://doi.org/10.1016/j.ijforecast.2022.11.005

Zhang, G. P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing, 50,* 159-175. https://doi.org/10.1016/s0925-2312(01)00702-0