International Academic Institute for Science and Technology

International Academic Journal of Humanities
Vol. 13, No. 3, 2026, pp. 35-42.

ISSN 2454-2245

International Academic Journal of Humanities

www.iaiest.com

Predictive Analytics in Economics: Data-Driven Approaches for Market Forecasting and Policy Planning

Dr. Arasuraja Ganesan [a*]

[a*] *Associate Professor, Department of Management Studies, St. Joseph's College of Engineering, OMR, Chennai, Tamil Nadu, India. E-mail: arasuraja.mba@gmail.com*

Abstract

The economics world is changing from a crisis that explains a paradigm to a paradigm for shocks to come. The amount of high-frequency and alternative data, coupled with low-cost computing, has transformed predictive analytics into a tool that can be used for market forecasting and policy planning. This conceptual paper brings together recent literature on data-driven analytics and provides an overarching framework for data-driven analytics that links data ingestion, feature engineering, hybrid econometric–machine learning modeling, calibration, and decision support. It will explore the ways in which new indicators help to broaden the economic toolkit, the ways in which hybrid models bridge the gap between interpretability and accuracy, and the ways in which forecasting relates to market and policy applications. It also includes the theme of "structural breaks", algorithmic feedback loops, and opacity as limitations on foresight. The paper ends with some recommendations both for policymakers and market players and an agenda for further research.

Keywords: Predictive Analytics, Economic Forecasting, Machine Learning, Econometrics, Policy Planning, Big Data.

DOI: 10.71086/IAJH/V13I3/IAJH1316
Received: 11 June 2026/Revised: 22 July 2026/Accepted: 21 August 2026/Published: 30 September 2026

The Quantitative Turn: From Retrospective History to Prospective Foresight

Breaking the Mirror

Until very recently economics was a mirror and it foretold crises once had happened. Recessions, currency failures, and credit crises were recreated via after-the-fact studies, and the insights were passed on as theory. That is nothing but a move towards a prospective one where the target is to predict shocks in the market and to experiment with policy in advance. Patterns are being driven by this type of analytics, which can learn patterns from large and diverse data, and some recent work describes analytics based on this AI as a transformation of knowledge, innovation, and efficiency in data-rich contexts (Zong & Guan, 2025). This applies to any economic judgment or decision, as with any determination, the relevance and accuracy of the estimate are what are important.

The Big Data Deluge

Economists are seeing a change in the amount, speed, and type of data that is available to them. Along with quarterly and monthly statistics, information from the financial feeds, transactions, Web activities and sensor streams is now supplementing that information. There are reports of performance gains in organizations based on the use of big data analytics when it becomes part of management practice (Prakash, 2024). Predictive analytics also plays a part in commercial strategy studies, identifying opportunities, and improving profitability and business growth (Joel & Oguanobi, 2024). Besides firms, a rise in interest in machine learning and data-driven methods for the predictive analysis of complex technical-economic systems is noted: reviews point out the potential, as well as the unsolved problems (Strielkowski et al., 2023). AI-driven predictive analytics has been used in the context of a supply chain to provide insight into optimizing planning decisions based on data, closely related to aspects of demand, cost, and price for the chain (Iseri et al., 2025). These strands collectively demonstrate the impact of the growth of computational power and the abundance of data in altering the extent, timeliness, and granularity of what can be forecasted.

Scholarly Trajectory and Objectives

This paper will go over three questions. The first question is: What is the role of data-driven methods in the context of classical econometric methods in forecasting? Second, what are the architectural options that are sufficiently robust, such that can be used to the market and policy? Thirdly, what epistemic and ethical limits are there for use? This paper is a conceptual and integrative paper. It does not provide any new primary experiments, but rather it combines the experiments from the reference list, and it builds an outline. The theoretical background is discussed in Section II, the architecture is described in Section III, empirical implementations are described in Section IV, and conclusions are presented in Section VI.

The contribution is threefold: an integrated pipeline with econometric and learning components embedded in one pipeline; a mapping of existing evidence to tasks in the market and policy; and a clear method for testing the performance of the forecasts based on the data vs the institutional forecasts.

Epistemic Convergence: Bridging Classical Econometrics and Machine Learning

The Two Cultures of Modeling

Until now, traditional economic modeling logic has been to first stipulate the relationships between variables in a model, then estimate the parameters of the model with data. In contrast, data-driven algorithms give the flexible functional forms the data to learn from; and the success measure is accuracy on out-of-sample data. In modeling methodologies employed in energy systems are discussed, and three types of models are identified – data-driven, mechanism-driven and hybrid, with the caveat that a hybrid approach can mitigate the drawbacks of either approach (Lin et al., 2025). The same is true in economics: in the economics world, structural models replace mechanism-driven culture, and the data-driven one is represented by machine learning. The data-driven side is found to be a valuable aspect of the studies through applied work (Bachmann et al., 2022). For example, to address questions fundamental to fiscal and environmental policy, a “temporal convolutional network” has been applied to directly predict outcomes of carbon tax policies from data, showing that learned models can offer answers to the questions (Chen et al., 2024).

It is important to state here the distinction between "forecasting" and "inference". Causal effect and good effect identification are not necessarily correlated with a good model. A good model is not necessarily good effect identification, and good effect identification is not necessarily a good model out of sample. The two cultures are thought of as complementary, and the two learning methods are tailored to the two types of tasks – prediction, pattern discovery; and structural, counterfactual reasoning and policy evaluation.

Expanding the Economic Sensorium

The other change in the indicators is that of the indicators themselves. For a small set of variables, the information was provided using Official statistics in classical forecasting. The data available is now also based on text-based sentiment analysis, satellite data on economic activity, and mobility data from mobile devices and, above all, is available in advance of official releases. An exogenous variable included in a seasonal ARIMA model (SARIMAX) using the web search data increased the forecasting accuracy of the tourism demand, which is an improvement over a traditional time-series model with behavioral signals (Lee, 2025). Similarly, empirical research on food supply chains reported that big data and predictive analytics were able to improve the social sustainability performance, which suggests that predictive capability can have an impact on social sustainability beyond the limited financial metric (Bag et al., 2024). These are signals that extend the 'economic sensorium' but also contain noise, revisions, and selection effects that need to be dealt with.

Methodological Tensions

There is an acceptable level of interpretation and accuracy. Causal statements and counterfactuals are bolstered by structural models, but may have poor forecasts when there are changes in relationships; flexible learners may have good forecasts but do not explain the causal mechanisms. In the majority of instances, when forecasts are used in decisions for the public, most important. However, the benefits of these technologies depend on how are used and responding to them, as well as the quality of technology performance, which is reflected in the work on data driven technologies and SDGs.The advantage of such technologies, however, is not only on technology performance but also on the way how applied and managed, as highlighted in the work done on data-driven technologies and SDGs. The hybrid solution is proposed in Lin et al., (2025) - learning for the remaining structure and keeping the theory for the reliable aspects.

Table 1: Comparison of Econometric, Machine Learning and Hybrid Forecasting Approaches

Dimension	Econometric models	Machine learning models	Hybrid models
Functional form	Specified by theory; linear or parametric	Learned flexibly from the data	Theory-based anchor plus learned correction
Interpretability	High; coefficients carry economic meaning	Low to moderate; needs post-hoc explanation	Moderate; the anchor remains interpretable
Stable regime	Reliable, but may miss non-linearities	Often strong accuracy	Usually, strong accuracy
After a structural break	Degrades slowly; theory restrains drift	May degrade sharply when patterns change	Degrades less if the anchor remains valid
Best suited to	Inference and counterfactuals	Nowcasting and pattern discovery	Forecasting that informs policy

The differences between the two modelling cultures explained in Section II and the hybrid designs offered in Section III are summarized in table 1. The rows are related to functional form, interpretability, data, behaviour in stable regime vs broken regime and suitability. Econometric models are based on relationships as stated in the theory - the coefficients have an economic interpretation, and they can be used to do counterfactual reasoning; however, the models may not capture non-linear patterns and may only react slowly to new observations. A machine learning model can learn the flexible form from large volumes of high-frequency and alternative data, but is not necessarily great at learning in stable periods and can fail very quickly when the "stable period" that it learned changes. Along with theory-based anchor, a learned correction: Hybrid models. The table is qualitative, and summarizes the argument of the literature reviewed, not new estimates. It will provide guidance to practitioners on the choice of design for a particular task, e.g. Nowcasting, pattern discovery and forecasting (hybrids to guide policy). It also explicitly makes the accuracy vs. interpretability trade-off, which allows the trade to be documented and subsequently audited.

Engineering the Oracle: Architectural Blueprints of Predictive Economic Models

Structural Pipeline Anatomy

A predictive economic system is most effectively thought of as a pipeline – or an ensemble of models. Fig. 1 presents the 5 stages. Market and macroeconomic data, and alternative data, are ingested in a way that matches frequency and eliminates outliers and gaps from the data sets. Feature engineering then creates lags and volatility proxies, along with text-based indicators. The modeling layer parallelizes the econometric, machine learning, and deep learning components. The outputs of these stages are merged and checked in a fusion and checking stage, and the fused and checked outputs of these stages are presented as market forecasts and policy scenarios in a final stage. It is the

actualization of results that goes hand in hand with data refresh and re-training of models – closing the loop. This lifecycle view is comparable to the applied research recently developed Zong & Guan, (2025) and Iseri et al., (2025) that provides an end-to-end view of analytics.

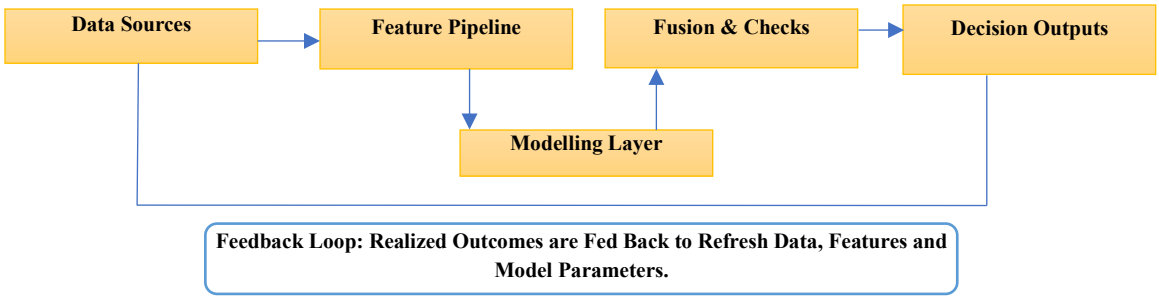


Fig. 1: Integrated Econometric and AI Framework for Market Forecasting and Policy Analysis

Fig. 1 shows an integrated forecasting model, which integrates market feeds, macroeconomic statistics, and alternative data. The feature pipeline has five stages: ingestion, cleaning, imputation, frequency alignment, and feature engineering. The modeling layer handles patterns in the complementarity, using econometric methods, machine learning ensembles, and deep learning architectures. The predictions are evaluated and updated via model stacking, calibration, walk-forward validation, and structural break stress tests. The decision output includes predictions of volatility, demand, and credit risk, along with scenarios for fiscal and monetary policy. A feedback loop returns the actual results back into the system to refresh data, features, and model parameters, which can allow changes to the model as the market changes and new data becomes available over time.

Algorithmic Fusion

The hybrid models combine both the econometric and the learning models. The first design is where a time-series model (e.g. ARIMA or GARCH) is used to model trend, seasonality and conditional volatility, and a learning model is fit to the residuals of this time-series model or other covariates. The simplest case is the SARIMAX specification of Lee, (2025), which adds exogenous search information to the classic case. The second approach, related to deep sequence models, learns non-linear dynamics directly (Chen et al., 2024). In the third, called ensemble methods, a number of forecasters are combined, and the errors cancel out. A general conclusion in Lin et al., (2025) about hybrid approaches also supports the notion that a hybrid of the extremes is more likely to succeed than either extreme alone.

A stylized hybrid forecaster is a forecaster in which each of the k forecasters is an econometric model, a machine learning model or a deep learning model that uses the feature vector x_t, and the weights w_k are positive and sum to one and are estimated from recent out-of-sample forecasting performance.

A residual-learning variant is equation (1)

$$\dot{y}(t+h) = \dot{y}_E(t+h) + g(x_t, e_t) \qquad (1)$$

Where $\dot{y}_E$ is the econometric forecast and g is a function to be learned to correct the error in the econometric forecast, e_t. In both forms, there is an economically meaningful anchor which is not changed by the data.

Calibration and Robustness Protocols

Overfitting and structural breaks are the most common causes of failure of predictive systems. Therefore, there are three protocols that are recommended. Walk-forward (rolling-origin) validation provides for the time ordering, and it is not prone to look-ahead bias. Fitting noise is prevented by regularisation, early stopping, and simple benchmark modeling. Stress tests reestimate/re-evaluate models on the turbulent sub-periods to ensure that accuracy is not lost due to changes in the regime. Forecast intervals and calibration checks should be given as well as a point forecast, enabling end users to appreciate the uncertainty of a point forecast. Finally, institutional treatment of risk is required - each forecaster using a model needs to have a name, a training period, a monitoring period, and a retirement rule; models that get it wrong should be retired promptly, not by default.

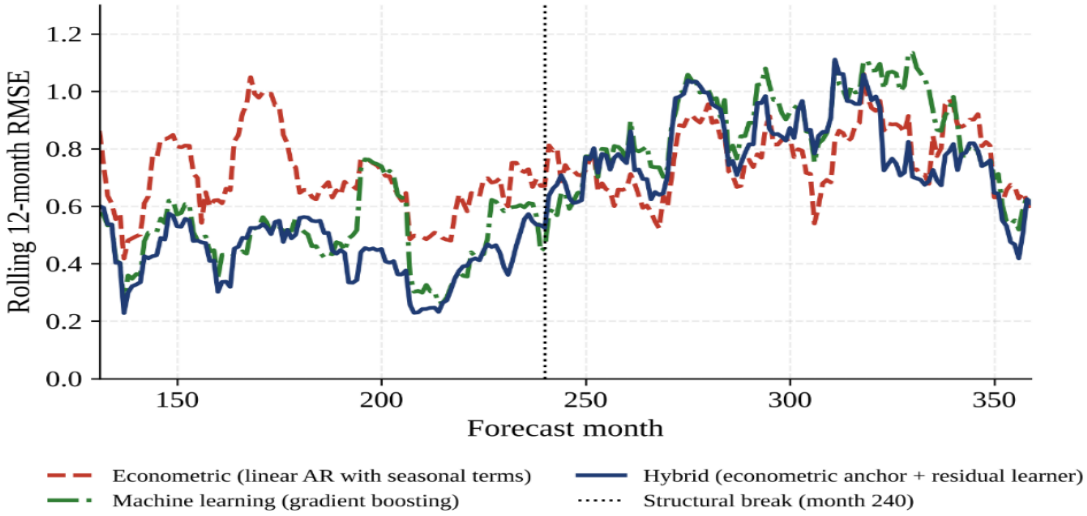


Fig. 2: Rolling Twelve-Month Forecast Error of Econometric, Machine Learning and Hybrid Forecasters around a Structural Break

The RMSE for the three forecasters on the synthetic series over the last 12 months is presented in fig. 2. The horizontal axis is the subsequent month that you are predicting, and the vertical axis is the error; the first time you'll be better, the lower the error line will be. The structural break at month 240 is indicated by the dotted vertical line. The non-linear exogenous variable is not captured by the econometric forecaster, and prior to the break, the econometric forecaster has the biggest errors whereas the machine learning and hybrid forecasters follow the series closely. Following the break, each line will increase in height and lines will intertwine since the relationships learned from the calm data no longer hold true. The worst rolling error after the break is 1.13 for the machine learning forecaster, but it is about 1.05 for the econometric forecaster. The hybrid forecaster has the lowest mean absolute error of 0.50 throughout the evaluation period, while the FNN and LNN models have mean absolute errors of 0.56 and 0.62, respectively, and the highest mean directional accuracy of 76 percent, compared to 75 percent for the FNN and 70 percent for the LNN models, respectively. The figure is a key part of the argument of the paper that flexible learners learn during calm times, are fragile during more turbulent times and hence forecast accuracy needs to be rechecked frequently, and there needs to be drift detection and stress testing carried out on turbulent sub-calm times. The data are 'made-up.

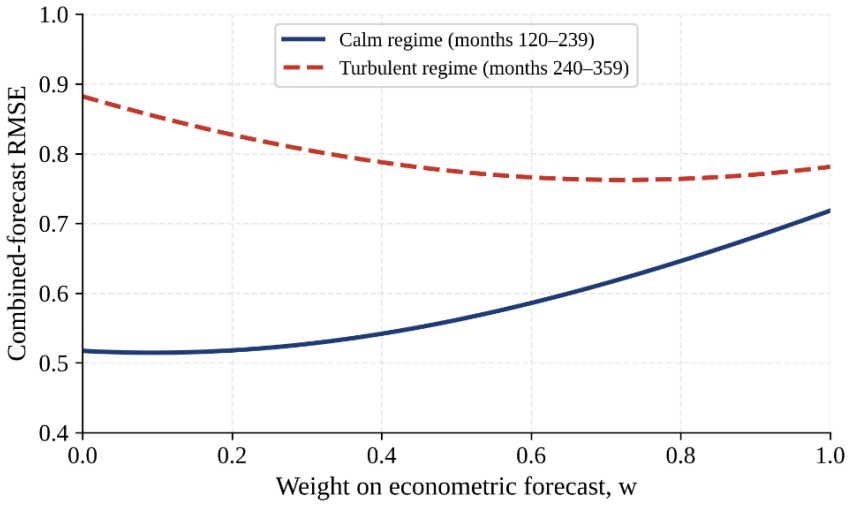


Fig. 3: Error of the Combined Forecast as a Function of the Weight on the Econometric Forecaster in Calm and Turbulent Regimes

The RMSE of the combined forecast of Equation (2) is plotted in Figure 3 as a function of the weight attached to the econometric forecaster in the two regimes, calm and turbulent. The horizontal axis runs from 0 (all machine learning forecast) to 1 (all econometric forecast) and the vertical axis is the error. The curves are different and the point is that they are different. In the calm regime the error is lower around w = 0.09, at 0.515, and increases consistently with the increasing weight being given to the econometric forecast, so that most weight should be placed on flexible learning. The reverse is true for the turbulent regime. The error of the pure machine learning forecast is 0.88, which is not far off from the pure econometric forecast error of 0.78, and the minimum forecast error at an interior weight of 0.72 is 0.76. The blend works best at the extremes when the conditions are most challenging, as in the hybrid designs studied by Lin et al. (2025). In practice the combination weights shouldn't be fixed. They should be estimated as the stylised hybrid forecaster argued - from the most recent out-of-sample performance - and should be checked for drift, so that the econometric anchor would be given even more weight after a lull.

Market Pulses and Policy Stress-Testing: Empirical Implementations

Anticipating Market Turbulence

This section presents a comparison between the empirical evidence identified in the literature examined and the three areas of interest to the application: asset volatility, consumer behavior, and credit risk. It does NOT make new estimates. There is a relationship between each study and economic prediction. The best representation of consumers' behavior is demanding forecasting and is applied in the case of applying search data to predict tourism demand. Predictive analytics' influence on business and supply chain studies can be seen in terms of expansion, profitability, and planning decisions. The framework uses existing econometric models that are already in use (GARCH for credit volatility, classification learners for credit default) and are implemented in the corpus reviewed.

The evidence is gathered in applied fields where results are easily observed, and thus validation is easy. In contrast, macro-financial applications deal with uncommon events and rather sluggish variables, resulting in significantly fewer test observations, and for these applications claims of predictive performance should be made with noticeable caution.

Proactive Macro-Policy Simulation

Stress testing fiscal and monetary interventions can be done with forward-looking forecasts and/or scenario analysis. The principle is illustrated in the carbon tax projection, where a projection of a policy-relevant quantity is calculated by a learned model and then can be recalculated with different assumptions. The use of a policy simulation would also require the use of forecasts of inflation, output, and financial conditions, and then report on the distribution of results, and not a single path, as a simulation of the policy rate or tax rate would be applied to this. Here, the preferred model is the hybrid model, as it maintains structural coherence across scenarios. The pipeline approach also has an upside: the actual trajectory can be used to fine-tune scenarios as needed, and the gap between simulated and realized trajectories can be used to test the model's suitability.

Comparative Benchmark Analysis

The question is not empirical and cannot be answered or assumed: Are data-driven frameworks more predictive than the traditional tools central banks use? It is recommended that candidate models be compared using the same information sets and horizons as an institutional benchmark, that the same measures are used (root-mean-square error, mean absolute error, and directional accuracy), and that the Diebold–Mariano test is used to assess significance. It is recommended that performance results be reported separately for calm and turbulent periods. Reviews of the literature indicate that short-horizon nowcasting benefits from a variety of alternative data, and that structural models remain useful for long-horizon and counterfactual forecasting. These are NOT findings, but rather hypotheses to be tested in the future.

The Limits of Sight: Epistemic Blind Spots, Bias, and Systemic Risk

All statistical predictions rest on the concept that the future is like the past in the matter to which it is applied. That's not the case if new shocks occur that were not experienced in the past, and if a model is learned from past history, it has not seen a new regime. The review of these methods notes that data-driven methods still face unsolved problems under changing conditions. The smart thing to do is to be humble in design: Apply scenario analysis in addition to forecasting, look out for regime change, and maintain human judgment. Besides rare catastrophic events, the accuracy meter can be fooled by evolutionary shifts in the structure of markets, consumer preferences, or regulations. In other words, rolling re-estimation and drift detection are not critical events; they happen regularly and routinely.

When machine predictions shape public policy, it becomes a governance issue. Any officials who make a prediction should be able to explain the prediction and be held accountable if there is an error. The demand for transparency is backed up both by the governance in which technology benefits rely on, and partly by a preference for partly mechanistic models. Governance implications include a need for transparency, and partly mechanistic models are preferred in order to benefit from technology. In addition, there is a further risk of bias in the training data, as a model can perpetuate the inequalities in credit and welfare decisions history, and the social benefits of applied work should not be assumed as equal.

Strategic Synthesis and Future Trajectories

Predictive analytics doesn't supplant economic theory; it is a step toward anticipation. The data-driven approaches give better indicators, flexible functional forms, and fast updates, and the hybridization designs give a balance between the accuracy and the structure from the literature reviewed. At the same time, indeterminacy, self-reflexivity, and opacity are part of the limits of the possibility of prediction. The key thing to remember is that predictive analytics should judge its success in terms of decision-making, not on its predictions. Five directives follow. Develop real-time data infrastructure, leverage alternative data in conjunction with official data. Secondly, employ hybrid and ensemble models and follow up validated procedures. Third, make the forecast uncertainty and benchmark results public for audit purposes to allow claims of superiority to be verified. Fourth, establish algorithmic forecast governance, such as transparency, bias checks, and human control. Fifth, spend on skills: Forecasting teams need to spend some money on economists with an understanding of algorithms, and some on data scientists with an understanding of economic mechanisms, because there are hybrid models in between, such as those promoted in this article. Facilities with only one side involved will not be able to comprehend, pose questions about, or improve the model. There are a number of directions that should be noted. Rich optimization and simulation in economic models could be realized in the future in quantum computing. Using real-time agent-based macroeconomic simulations, calibrated from real-time data, may be able to incorporate feedback effects which are not represented in aggregate models. Causal machine learning could help resolve the interpretability problem and empirical research should assess the benchmark hypotheses in Section IV on actual forecasting data. In such studies, it would be useful to have common datasets and open evaluation code; this would enable independent researchers to test claims of predictive superiority and enable comparisons between architectures on the same data, under the same conditions.

References

Zong, Z., & Guan, Y. (2025). AI-driven intelligent data analytics and predictive analysis in Industry 4.0: Transforming knowledge, innovation, and efficiency. *Journal of the knowledge economy*, *16*(1), 864-903. https://doi.org/10.1007/s13132-024-02001-z

Prakash, D. (2024). Data-driven management: The impact of big data analytics on organizational performance. *International Journal for Global Academic & Scientific Research*, *3*(2), 12-23. https://doi.org/10.55938/ijgasr.v3i2.74

Joel, O. T., & Oguanobi, V. U. (2024). Data-driven strategies for business expansion: Utilizing predictive analytics for enhanced profitability and opportunity identification. *International Journal of Frontiers in Engineering and Technology Research*, *6*(02), 071-081. https://doi.org/10.53294/ijfetr.2024.6.2.0035

Strielkowski, W., Vlasov, A., Selivanov, K., Muraviev, K., & Shakhnov, V. (2023). Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review. *Energies*, *16*(10), 4025. https://doi.org/10.3390/en16104025

Iseri, F., Iseri, H., Chrisandina, N. J., Iakovou, E., & Pistikopoulos, E. N. (2025). AI-based predictive analytics for enhancing data-driven supply chain optimization. *Journal of Global Optimization*, 1-28. https://doi.org/10.1007/s10898-025-01509-1

Bachmann, N., Tripathi, S., Brunner, M., & Jodlbauer, H. (2022). The contribution of data-driven technologies in achieving the sustainable development goals. *Sustainability*, *14*(5), 2497. https://doi.org/10.3390/su14052497

Lin, Y., Tang, J., Guo, J., Wu, S., & Li, Z. (2025). Advancing AI-enabled techniques in energy system modeling: a review of data-driven, mechanism-driven, and hybrid modeling approaches. *Energies*, *18*(4), 845. https://doi.org/10.3390/en18040845

Chen, J., Cui, Y., Zhang, X., Yang, J., & Zhou, M. (2024). Temporal convolutional network for carbon tax projection: A data-driven approach. *Applied Sciences*, *14*(20), 9213. https://doi.org/10.3390/app14209213

Lee, G. C. (2025). A data-driven approach to tourism demand forecasting: Integrating web search data into a SARIMAX model. *Data*, *10*(5), 73. https://doi.org/10.3390/data10050073

Bag, S., Srivastava, G., Cherrafi, A., Ali, A., & Singh, R. K. (2024). Data-driven insights for circular and sustainable food supply chains: An empirical exploration of big data and predictive analytics in enhancing social sustainability performance. *Business Strategy and the Environment*, *33*(2), 1369-1396. https://doi.org/10.1002/bse.3554