PREDICTING BUSINESS GROWTH:

A REVIEW OF BEST PRACTICE ECONOMETRIC AND MACHINE LEARNING APPROACHES

Dr Rita Nana-Cheraa, Dr Michalis Papazoglou and Prof Stephen Roper

Warwick Business School, Oxford Brookes University, Warwick Business School

January 2026

Executive Summary

The difficulty of forecasting business growth has long engaged economists and scholars in business and management. Yet, empirical research has found it hard to pinpoint consistent growth drivers, with most models showing very low predictive accuracy. This unpredictability stems from the ongoing heterogeneity of firms and the fact that variations across industries, technologies, and countries make generalisation challenging.

While some stylised facts exist, such as the tendency for younger and smaller firms to grow faster, conventional econometrics models often fail to explain future growth. ML techniques, however, offer new opportunities by processing high-dimensional and unstructured data sources (e.g., financial reports, web content) to uncover hidden relationships. In this review, we provide an accessible overview of the latest developments in modelling business performance.

Econometric models

Econometric models typically use a deductive approach, guided by theoretical or conceptual frameworks that develop testable hypotheses about what affects business growth. Most studies look at various growth indicators, including employment, sales, productivity, assets, exports, and profitability. Employment growth is the most frequently analysed metric in the 19 studies considered here (12 studies), followed by sales and labour productivity growth (6 studies each), and total factor productivity (TFP) and asset growth (3 studies each).

Growth predictors encompass a broad range of areas: human and knowledge capital, innovation, R&D, support, leadership and governance structures, financial resources and access to credit, market conditions, institutional environments, and policy and regulatory frameworks.

Although OLS and panel models are the most common methodological approaches, they generally exhibit lower predictive power than alternative methods. For example, OLS estimations often yield R-squared values below 0.09. Panel estimations show greater variation, with R-squared values ranging from 0.026 and adjusted R-squared values between 0.017 and 0.304, highlighting differences in explanatory power across various contexts and specifications.

When analysing individual studies, the Difference-in-Differences approach combined with Propensity Score Matching (PSM) provides the highest predictive power, with adjusted R-squared values ranging from 0.88 to 0.98 depending on the growth model specification. Quantile regression also demonstrates strong explanatory ability, with pseudo-R-squared values between 0.68 and 0.80.

Machine Learning (ML) and AI-based approaches

ML enables computers to learn from data and enhance their performance on specific tasks by recognising patterns and making predictions or decisions based on experience rather than fixed rules, with little or no human input and without explicit programming. ML algorithms generate predictions by searching data for complex associations between variables.

At a high level, ML is categorised into three main types: supervised learning, unsupervised learning, and reinforcement learning. While supervised learning relies on labelled data to predict outcomes, unsupervised learning detects hidden structures in unlabelled data, and reinforcement learning involves decision-making through feedback from an environment. Each category offers unique analytical and predictive capabilities that serve diverse applications, from financial forecasting to autonomous systems.

Supervised learning (SL) is the most popular ML method and involves training a model on labelled datasets to link input variables with known output variables (Maple et al., 2023). During this process, the model identifies patterns that allow it to predict future or unseen results accurately. SL has been used to forecast company performance, solvency, and overall success by pinpointing the most influential variables affecting outcomes.

For example, supervised ML algorithms have been used to predict which firms will achieve high growth alongside econometric approaches (i.e., Logistic Regression). When both approaches are employed, ML algorithms outperform econometric models in forecasting high-growth firms, demonstrating the predictive power of ML techniques.

Contrasting strengths

Both econometric models and ML techniques aim to learn from data, but they differ in philosophy and purpose. Econometric models, based on statistical theory and economic reasoning, are mainly used for hypothesis testing and causal inference. They rely on predefined theoretical frameworks and

assumptions about data distribution, emphasising interpretability and formal inference through confidence intervals and significance tests.

In contrast, ML methods are driven by algorithms and are less limited by theoretical assumptions. Their main aim is predictive accuracy rather than inference, focusing on improving performance through computational learning. While econometrics aims to confirm or refute predefined hypotheses, ML seeks to identify complex, often non-linear patterns in large datasets without relying on assumptions about data distributions or model structures.

Despite its advantages, ML's focus on predictive performance creates interpretability challenges often called the "black box" problem. (Huang et al., 2024; Valizade et al., 2024). Unlike econometric models, where coefficients give direct insights into the relationships between variables, ML algorithms usually offer limited transparency about how input features affect outcomes. This lack of clarity and interpretability raises concerns, especially for policymakers, managers, and investors, who care not only about prediction accuracy but also about the main factors driving a firm's potential for high growth.

Essentially, econometric and ML paradigms are complementary rather than mutually exclusive. As both fields evolve, a more integrated, boundary-expanding methodological paradigm is emerging, capable of balancing interpretability with predictive power and blending econometric rigour with ML flexibility to generate more robust, generalisable, and theoretically meaningful insights.

Practical implications

Implementing either econometric or ML approaches involves several specific choices related to the goals of the predictive task, data availability, and transparency. These issues are summarised in the following table:

Criterion	Econometric Approach	ML/AI Approach	Real-World Example
Primary Goal	Hypothesis testing, causal inference	Predictive accuracy, pattern recognition	Econometric: Assessing impact of R\&D grants on SME growth (e.g., Vanino et al., 2019). ML: Predicting high-growth firms using Random Forest (e.g., Houle & Macdonald, 2025).
Interpretability	High (coefficients, significance tests)	Low to medium (often "black box"; explainable AI needed)	Econometric: Quantile regression showing R\&D effects at different growth quantiles (Coad et al., 2016). ML: Neural networks predicting revenue growth but hard to interpret (Houle & Macdonald, 2025).
Data Requirements	Structured, longitudinal/panel data; smaller datasets	Large, high-dimensional, possibly unstructured data	Econometric: Longitudinal Small Business Survey (UK). ML: Web-scraped financial and social media data for firm success prediction.

Criterion	Econometric Approach	ML/AI Approach	Real-World Example
Assumptions	Strong (distributional, linearity, independence)	Minimal; non-parametric, flexible	Econometric: OLS models assuming linearity (Murro et al., 2023). ML: Gradient Boosted Trees handling non-linear interactions (Vuković et al., 2024).
Transparency	High (clear theoretical framework)	Lower (complex algorithms, harder to explain)	Econometric: DiD models for policy evaluation (Mulier & Samarin, 2021). ML: Deep learning for text-based growth prediction (Gangwani & Zhu, 2024).
Computational Demand	Low to moderate	High (requires significant computing resources)	Econometric: Panel regressions on survey data. ML: Neural networks trained on millions of observations.
Predictive Power	Generally low to moderate; better for causal insights	High for out-of-sample prediction	Econometric: R^2 often <0.1 for OLS models. ML: CatBoost achieving 86% accuracy for growth prediction (Vuković et al., 2024).
Theory Integration	Strong (based on economic reasoning)	Weak; primarily data-driven	Econometric: Testing Schumpeterian growth theory. ML: Inductive discovery of patterns without prior theory.
Handling Non-Linearity	Limited (requires transformations)	Strong (captures complex, non-linear relationships)	Econometric: Adding quadratic terms for size effects. ML: Random Forest capturing non-linear effects of age and leverage.
Adaptability to New Data	Limited; model structure fixed	High; models can retrain and adapt	Econometric: Static regression models. ML: Online learning algorithms updating predictions in real time.
Policy Usefulness	High (clear drivers of growth for policy design)	Lower (harder to justify decisions based on opaque models)	Econometric: Evaluating subsidy impacts for innovation policy. ML: Predicting which firms will become high-growth for investment targeting.
Sector-Specific Relevance	Strong if theory tailored	May require retraining for sector-specific patterns	Econometric: Sector-specific productivity models. ML: Industry-specific training for growth prediction in tech vs manufacturing.

Defining the aims of the predictive exercise is essential for selecting between ML and econometric approaches. ML methods may deliver superior predictive accuracy for a given dataset compared to purely econometric methods. However, all ML predictions are subject to the 'black box' problem, which means it may not be very clear how or why specific predictions are made. This complicates the use of these predictions to refine related policy initiatives or support measures. Conversely, econometric models— which establish a more explicit link between drivers and growth— offer more direct insights.

Other questions may also be important when examining growth within a specific group of businesses. In such cases, models trained on a broadly based database might be less relevant to particular sectors or firm size bands.

Predicting business growth with either an econometric or ML approach also demands substantial data resources, including growth metrics and potential explanatory or correlated variables for many companies, ideally spanning several years.

Finally, it is important to consider the transparency and persuasiveness of the two modelling approaches. ML methods may be seen as less transparent and possibly less reliable due to the 'black box' approach. Econometric methods may be more transparent but can also be challenging to communicate because of their complexity.

Now that you have read our report, we would love to know if our research has provided you with new insights, improved your processes, or inspired innovative solutions.

Please let us know how our research is making a difference by completing our short feedback form via this link.

You are also welcome to email us if you have any questions about this report or the work of the IRC generally: info@ircaucus.ac.uk

Thank you

The Innovation & Research Caucus

Authors

- Dr Rita Nana-Cheraa – Warwick Business School
- Dr Dr Michalis Papazoglou – Oxford Brookes University
- Prof Stephen Roper – Warwick Business School

Acknowledgements

This work was supported by Economic and Social Research Council (ESRC) grant ES/X010759/1 to the Innovation and Research Caucus (IRC) and was commissioned by Innovate UK. The interpretations and opinions within this report are those of the authors and may not reflect the policy positions of Innovate UK.

About the Innovation and Research Caucus

The IRC supports the use of robust evidence and insights in UKRI's strategies and investments, as well as undertaking a co-produced programme of research. Our members are leading academics from across the social sciences, other disciplines and sectors, who are engaged in different aspects of innovation and research system. We connect academic experts, UKRI, IUK and the ESRC, by providing research insights to inform policy and practice. Professor Tim Vorley and Professor Stephen Roper are Co-Directors. The IRC is funded by UKRI via the ESRC and IUK, grant number ES/X010759/1. The support of the funders is acknowledged. The views expressed in this piece are those of the authors and do not necessarily represent those of the funders.

About the Enterprise Research Council

The Enterprise Research Centre (ERC) is an independent research centre based at Warwick Business School focusing on growth, innovation and productivity in small and medium-sized enterprises. The Centre is funded by the Economic and Social Research Council, The Department for Business and Trade, The Department for Science Innovation and Technology, Innovate UK, the British Business Bank and the Intellectual Property Office. The views expressed in this report are those of the authors and do not necessarily represent those of the funders.

Find out more

Contact: info@ircaucus.ac.uk

Website: https://ircaucus.ac.uk/