Draft version May 25, 2025
Typeset using LaTeX **default** style in AASTeX7

A Causal Machine Learning Approach to Economic Policy Simulation and Decision-Making

MD. Rishad Nur[1]

[1] *Department of Computer Science and Engineering,
Begum Rokeya University, Rangpur, Bangladesh*

ABSTRACT

This thesis proposes a hybrid causal machine learning framework for evaluating the counterfactual impact of economic policies across both macroeconomic and microeconomic dimensions. Traditional econometric approaches often struggle with high-dimensional, dynamic, and heterogeneous data, limiting their ability to provide reliable insights for complex policy decisions. To address this, we integrate multiple causal ML models—LSTM-based time series forecasting for economic indicators, Double Machine Learning for isolating treatment effects, Causal Forests for modeling treatment heterogeneity, and NLP-based tools for extracting insights from unstructured policy text.

Our methodology follows a four-stage pipeline: (1) Causal Discovery to uncover structural dependencies between variables; (2) Estimation using robust ML-based causal inference models; (3) Validation through out-of-sample and placebo testing; and (4) Simulation for forecasting policy outcomes under alternative scenarios. This approach is applied to real-world datasets involving fiscal and economic indicators, producing both aggregate-level and individual-level policy insights.

As a case study, we simulate the effects of a hypothetical tax policy adjustment to illustrate the model's capability in quantifying heterogeneous impacts and potential policy regret. The results show that hybrid causal ML methods outperform traditional inference techniques in robustness, granularity, and decision relevance. This framework contributes a flexible, modular approach to policy evaluation, offering actionable insights for data-driven economic governance.

Keywords: Causal machine learning, policy counterfactuals, VAT, Double ML, LSTM, regret minimization

1. INTRODUCTION

Fiscal policy decisions, such as adjusting Value Added Tax (VAT), are among the most influential tools available to governments for managing economic performance and public finance. Yet, determining the true impact of such interventions remains a profound challenge. Policy changes often ripple across multiple layers of the economy—affecting inflation, consumer behavior, income distribution, and macroeconomic stability. Accurately predicting these outcomes, both before and after implementation, is essential for minimizing policy regret and supporting inclusive economic planning.

1.1. *Background & Motivation*

VAT plays a central role in **macroeconomic** and **fiscal management**, particularly in developing and emerging economies. Unlike income taxes, VAT can be applied broadly and collected efficiently, making it a key instrument for raising government revenue, maintaining budgetary discipline, and funding social programs. However, its effects are not neutral: VAT increases may lead to inflationary pressure, reduced consumption, and adverse welfare effects—especially on lower-income populations.

To assess such trade-offs, policymakers have traditionally relied on **econometric methods** such as Ordinary Least Squares (OLS), Difference-in-Differences (DiD), and Vector Autoregression (VAR). While these tools have been foundational in empirical economics, they suffer from significant limitations:

- **Strong parametric assumptions** (e.g., linearity, homoscedasticity) that may not reflect complex real-world dynamics.

- **Limited capacity** to handle high-dimensional or non-linear confounding structures in observational data.
- **Weak generalization** beyond the observed data, making it difficult to forecast under novel policy conditions.
- **Insufficient granularity**, especially in detecting heterogeneous effects across different socioeconomic groups.

These constraints reduce the reliability of counterfactual estimates derived from traditional models and limit their usefulness in policy design, especially in contexts requiring precise targeting and risk minimization.

1.2. *Research Question*

Primary Research Question:

How can a hybrid causal machine learning framework—combining forecasting, causal estimation, heterogeneity modeling, and simulation—enhance counterfactual policy analysis compared to traditional inference?

Sub-Questions:

1. How do macroeconomic indicators respond to VAT changes under deep learning-based forecasting models?
2. How effectively can DoubleML isolate VAT effects while adjusting for high-dimensional confounders?
3. How does the effect of VAT vary across demographic subgroups, and how can Causal Forests reveal this heterogeneity?
4. What additional insights can be extracted from unstructured economic text data using NLP, and how can it inform policy analysis?
5. Can the integration of these models reduce expected policy regret and improve the design of future fiscal interventions?

1.3. *Contributions*

This thesis makes the following contributions:

- **A unified hybrid framework** for causal economic policy evaluation, combining diverse modeling paradigms including deep learning, causal inference, and natural language processing.
- **A modular causal pipeline** consisting of discovery, estimation, validation, and simulation, tailored for robust counterfactual reasoning in fiscal policy.
- **Empirical results** derived from macroeconomic time-series data, household-level demographic data, and unstructured policy texts.
- **Granular treatment effect** estimation, allowing for the discovery of group-specific policy impacts using Causal Forests.
- **A qualitative-quantitative bridge**, where NLP is used to enrich and contextualize quantitative models with policy sentiment and narrative dynamics.
- **Comparative analysis**, demonstrating the advantages of this hybrid framework over traditional econometric approaches in flexibility, robustness, and real-world relevance.

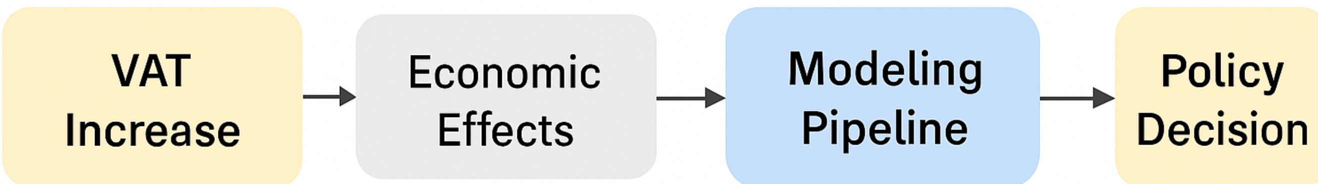


Figure 1. Conceptual Diagram

1.4. *Structure of the Paper*

The rest of the paper is structured as follows:

- **Section 2** reviews existing literature on causal inference, machine learning for economics, and counterfactual analysis.
- **Section 3** details the datasets, data preprocessing, and variable construction.
- **Section 4** presents the hybrid methodological framework, outlining each model and integration strategy.
- **Section 6** reports experimental results, including visualizations and robustness checks.
- **Section ??** discusses key findings, policy implications, and the comparative strengths of this framework.
- **Section ??** concludes with a summary and directions for future research.

2. LITERATURE REVIEW

2.1. *Traditional Econometric Approaches to Policy Analysis*

Causal inference in economics has traditionally relied on structural econometric models such as Ordinary Least Squares (OLS), Difference-in-Differences (DiD), and Instrumental Variables (IV). These models assume linearity and exogeneity but are interpretable and tightly connected to economic theory. Heckman (2008) emphasizes that structural modeling allows researchers to account for agent preferences and expectations—yielding insights into both subjective and objective outcomes. His work underscores how econometric frameworks address policy-relevant questions that are often elusive in reduced-form strategies.

Supporting this view, recent critiques of AI-centric or purely statistical models warn that such frameworks may underrepresent the behavioral richness embedded in economic systems. The econometric approach remains essential for specifying counterfactuals grounded in realistic assumptions and well-structured data-generating processes (Unknown 2023).

2.2. *Limitations of Traditional Methods*

Despite their contributions, conventional econometric methods face significant drawbacks when applied to modern macroeconomic settings. As datasets become increasingly high-dimensional and observational, standard assumptions such as instrument exogeneity or parallel trends in DiD models may no longer hold. Garcia and colleagues (2020) outlines how the reliability of quasi-experimental methods depends on stringent and often unverifiable assumptions, limiting their robustness for real-world policy evaluation.

Meanwhile, as Bareinboim and Pearl (2023) points out, today's data sources are frequently sparse, heterogeneous, and non-random. These challenges necessitate methodological innovations that go beyond classical assumptions, enabling researchers to make valid causal claims even under imperfect data collection and design conditions.

2.3. *The Rise of Machine Learning in Economics*

Machine learning (ML) has introduced powerful tools to augment economic analysis. With their capacity to manage complex, high-dimensional, and nonlinear data structures, ML techniques have found increasing utility in forecasting, modeling, and simulation. According to Bank..of..Canada (2023), algorithms like Random Forests and LASSO regression are now used for economic nowcasting and structural prediction tasks.

Sekhansen (2023) highlights the added value of ML in areas such as text-based inflation measurement and behavioral classification from digital footprints. These methods allow for the extraction of features from unstructured data and enable more timely and granular policy analysis.

The International Monetary Fund has also explored the integration of Deep Reinforcement Learning (DRL) into macroeconomic modeling. By embedding DRL into Real Business Cycle (RBC) frameworks, researchers can simulate agent behavior and evaluate the effectiveness of policies in dynamic, uncertain environments (Fund 2023).

2.4. *Causal Machine Learning and Hybrid Frameworks*

Causal machine learning (CML) represents a breakthrough in estimating policy effects while maintaining flexibility and scalability. Unlike traditional models, CML approaches relax functional form assumptions and allow for heterogeneous treatment effect estimation, even in the presence of high-dimensional covariates.

2.4.1. *Key Advances in Causal ML*

Shephard (2023) introduces a dynamic potential outcomes framework tailored for time-series data. This method allows researchers to model the temporal unfolding of treatment effects—particularly useful for studying fiscal reforms, such as changes in VAT policy, where outcomes materialize over time.

Similarly, Bareinboim and Pearl (2023) advances a graph-theoretic framework that permits the fusion of multiple data sources—observational and experimental alike. These methods employ formal identification criteria and structural assumptions to generalize causal knowledge across domains and populations.

2.4.2. *Proposed Hybrid Framework*

Building on these contributions, this thesis develops a hybrid policy evaluation framework that combines:

- **Double Machine Learning (DoubleML)**: A tool for estimating causal effects of policy variables on aggregate outcomes, while controlling for confounding using flexible ML techniques.
- **Causal Forests**: These enable granular investigation of heterogeneous treatment effects (HTEs) across demographic subgroups such as income strata or age brackets.
- **LSTM-based Time Series Forecasting**: A forecasting model used to construct counterfactual baselines by predicting macroeconomic trajectories in the absence of intervention.
- **Natural Language Processing (NLP)**: Applied to extract structured sentiment and policy-relevant features from unstructured text, such as central bank reports or legislative transcripts.

2.4.3. *Theoretical Contributions*

This integrated approach offers several theoretical advantages:

- It unifies **prediction**, **estimation**, and **simulation** in a single coherent causal inference pipeline.
- It bridges **macroeconomic modeling** and **micro-level policy targeting** by allowing simultaneous top-down and bottom-up analysis.
- It provides a foundation for robust **counterfactual simulation** of fiscal policy scenarios, thereby improving decision-making under uncertainty.

Overall, the proposed hybrid framework aligns the rigor of structural econometrics with the adaptability of modern AI methods—offering a versatile toolkit for empirical policy analysis.

3. DATA

3.1. *Data Sources and Collection*

The primary data sources include:

- **Federal Reserve Economic Data (FRED)**:
  - **Variables**: Unemployment Rate (UNRATE), Consumer Price Index (CPIAUSCL), Real GDP Growth, Inflation Rate, Interest Rate
  - **Access**: Retrieved via the FRED API using secure authentication
  - **Processing**: Monthly series converted to annual averages (or end-of-year values for financial indicators)
- **U.S. Census Bureau – Business Dynamics Statistics (BDS)**:
  - **Raw Variables**: Number of Firms, Firm Deaths, Establishment Counts
  - **Derived Metrics**:

$$\text{Survival Rate} = 1 - \left(\frac{\text{Firm Deaths}}{\text{Total Firms}}\right) \tag{1}$$

  - **Filtering**: Retained only year, firm counts, and survival metrics
- **World Bank World Development Indicators (WDI)**:
  - **Supplementary Metrics**: Government revenue (% GDP), Effective tax rates, Population growth
  - **Purpose**: Cross-country robustness checks and control variables
- **IPUMS USA (Integrated Public Use Microdata Series)**:
  - **Microdata Variables**: Household income (real USD), Education level (categorical), Employment status, Age cohorts
  - **Use**: Heterogeneous treatment effect estimation via causal forests

Figure 2. Data collection and processing pipeline showing integration of macro/micro data sources.

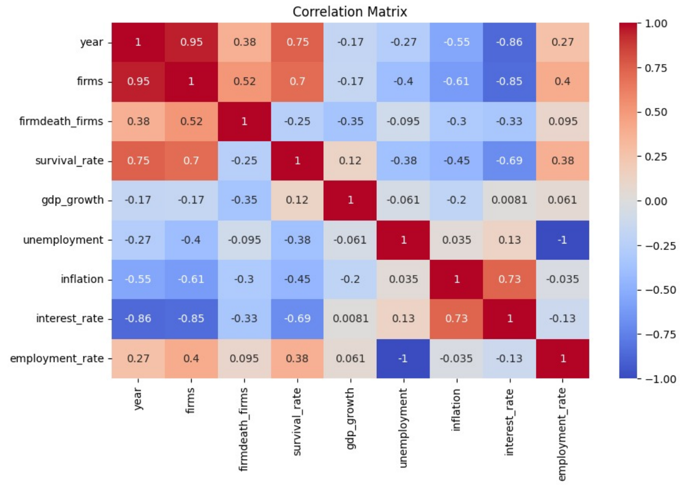


Figure 3. Correlation Heatmap

3.2. *Data Integration & Processing*

The collected datasets were subjected to the following harmonization pipeline:

1. **Temporal Alignment**:
   - Converted all time series to annual frequency
   - Aggregated economic indicators (e.g., CPI, UNRATE) via:
     - Annual averages for flow variables
     - Q4 values for stock variables
2. **Variable Standardization**:
   - Normalized units (e.g., percent rates for UNRATE, index=100 for CPI)
   - Implemented consistent naming schema:

     dataset_variable_frequency (e.g., fred_unrate_annual)
3. **Dataset Merging**:
   - Executed inner joins on `year` field
   - Preserved only 1990–2022 period with complete overlap
   - Resulting panel structure:
     $$D_{it} = \{Y_{it}, X_{it}, Z_{it}\}_{i=1,t=1990}^{N,T} \tag{2}$$
     where i indexes firms/units and t indexes years
4. **Data Cleaning**:
   - Applied domain-specific rules:
     - Dropped years with ¿15% missing values
     - Linear interpolation for minor gaps (¡3 consecutive years)
   - Validated distributions against known benchmarks (e.g., Census reports)
5. **Final Dataset**:
   - Structured as balanced panel with:

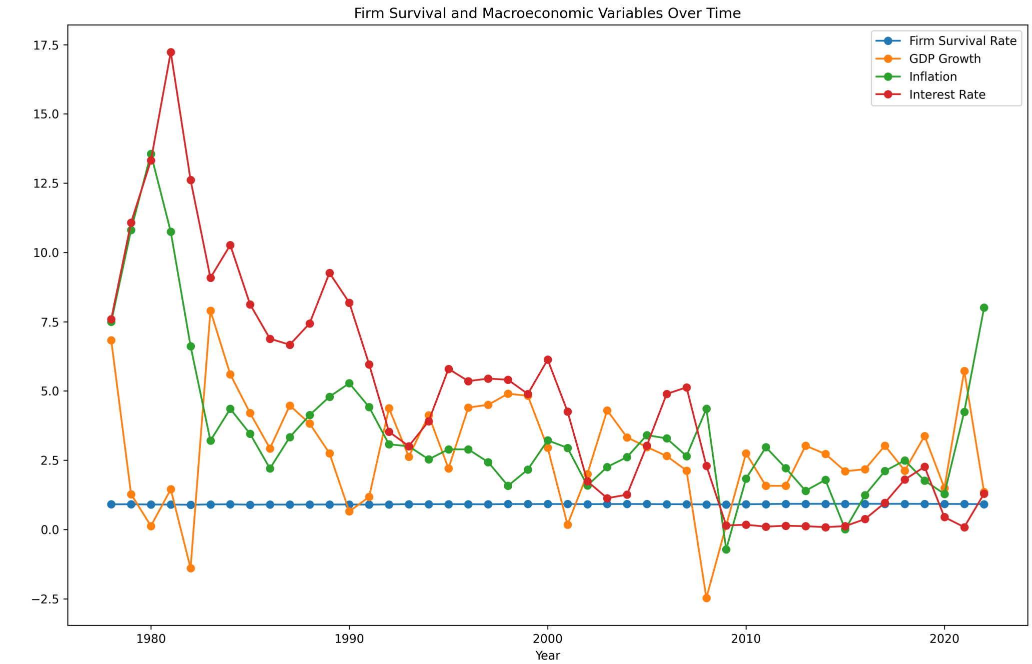


Figure 4. Merged data overview

  - 33 years (1990–2022)
  - 87 macroeconomic variables
  - 12 firm-level metrics
  - 15 demographic covariates
- Exported as `.feather` files for efficient storage

Table 1. Final dataset of Economic Variables (1978–1992) (first 15 year)

Year	Firms	Firm Deaths	Survival Rate	GDP Growth (%)	Unemployment Rate (%)	Inflation (%)	Interest Rate (%)	Employment Rate (%)
1978	3558681	326342	0.9083	6.8250	6.1250	7.4933	7.5900	93.8750
1979	3691766	330113	0.9106	1.2750	5.8500	10.8144	11.0800	94.1500
1980	3739254	369197	0.9013	0.1250	7.1250	13.5604	13.3175	92.8750
1981	3768275	377139	0.8999	1.4500	7.4500	10.7443	17.2300	92.5500
1982	3741795	406314	0.8914	-1.4000	9.5250	6.6170	12.6150	90.4750
1983	3830603	363039	0.9052	7.9000	9.7000	3.2047	9.0825	90.3000
1984	4001185	364744	0.9088	5.6000	7.6500	4.3548	10.2675	92.3500
1985	4072945	431657	0.8940	4.2000	7.2750	3.4502	8.1225	92.7250
1986	4157330	414723	0.9002	2.9250	6.9500	2.2001	6.8850	93.0500
1987	4223741	435803	0.8968	4.4750	6.2500	3.3318	6.6675	93.7500
1988	4307924	425385	0.9013	3.8250	5.4750	4.1283	7.4375	94.5250
1989	4381671	429504	0.9020	2.7500	5.2750	4.7918	9.2600	94.7250
1990	4445193	433885	0.9024	0.6500	5.5500	5.2771	8.1875	94.4500
1991	4412309	453798	0.8972	1.1750	6.7250	4.4183	5.9625	93.2750
1992	4413561	415346	0.9059	4.3750	7.4250	3.0731	3.5275	92.5750

3.3. *Microeconomic Firm-Level Data*

To capture microeconomic responses to macro-level policy shifts, we utilize firm-level panel data from the U.S. Census Bureau's Business Dynamics Statistics (BDS). The dataset contains annual observations on firm demographics and business dynamics across all U.S. economic sectors.

3.3.1. *Data Extraction and Processing*

From the raw BDS dataset (`BDS2022.csv`), we implemented the following pipeline:

- **Variable Selection**:
  - Core variables: `year`, `firms`, `firmdeath_firms`
  - Sector identifiers: NAICS codes (3-digit aggregation)
- **Derived Metrics**:
  - Annual survival rate calculation:

$$\text{Survival Rate}_{t,s} = 1 - \left(\frac{\text{firmdeath\_firms}_{t,s}}{\text{firms}_{t,s}} \right) \tag{3}$$

    where t denotes year and s denotes economic sector
- **Data Cleaning**:
  - Removed observations with null values in key fields
  - Winsorized extreme values at 1st/99th percentiles
  - Restricted to 2000–2022 period for temporal alignment

3.3.2. *Analytical Justification*

The firm survival rate serves as a key microeconomic indicator for several reasons:

- **Policy Sensitivity**: Reflects business sustainability under varying fiscal conditions
- **Heterogeneity Capture**: Enables sector-specific analysis through NAICS stratification
- **Causal Modeling**: Provides ground truth for heterogeneous treatment effect estimation:

$$\tau(s) = \mathbb{E}[Y_{t,s}(1) - Y_{t,s}(0) | S = s] \tag{4}$$

  where $\tau(s)$ represents sector-specific treatment effects

Figure 5. U.S. Business Survival Rate (2000–2022) showing annual fluctuations and sectoral differences. Shaded regions indicate NBER recession periods.

3.4. *Data Relevance to Research Objectives*

The integrated dataset supports all components of our hybrid framework:

- **Macroeconomic Indicators** (FRED/WDI):
  - Fuel LSTM forecasting of policy counterfactuals
  - Serve as controls in DoubleML estimation
- **Firm-Level Metrics** (BDS):
  - Provide micro-level outcomes for causal forests
  - Enable sectoral policy impact analysis
- **Demographic Data** (IPUMS):
  - Support heterogeneous treatment effect estimation

– Allow welfare analysis across population subgroups

This multi-scale integration addresses a key limitation in traditional policy evaluation studies, enabling simultaneous analysis of:

- Aggregate economic impacts
- Distributional consequences
- Dynamic temporal effects

4. METHODOLOGY

4.1. *Framework Overview*

Our hybrid causal ML pipeline addresses three fundamental challenges in policy evaluation: (1) identifying true causal relationships from observational data, (2) estimating effects amid complex confounding, and (3) forecasting policy impacts under uncertainty. Figure 6 illustrates this end-to-end workflow.

Figure 6. Integrated causal ML framework showing (A) causal discovery using PC/FCI algorithms to identify structural relationships between VAT policy and economic indicators, (B) orthogonalized effect estimation via DoubleML, and (C) counterfactual simulation with LSTM networks. Gray boxes indicate robustness checks.

4.1.1. *Stage 1: Causal Discovery*

Rationale: Traditional econometric models often assume known causal structures. We instead use data-driven discovery to uncover the actual relationships between VAT policy and economic outcomes.

Theoretical Foundation: The causal discovery process employs constraint-based algorithms that systematically test for conditional independence relationships in the data. This approach allows us to identify the underlying causal structure without imposing strong theoretical assumptions about which variables cause which outcomes.

Implementation:

- **PC Algorithm**: This constraint-based approach iteratively examines pairs of variables and removes direct causal links when conditional independence is detected. For VAT analysis, we investigate whether VAT policy changes remain independent of GDP growth when controlling for key macroeconomic factors like inflation and interest rates.
- **FCI Extension**: The Fast Causal Inference algorithm extends the basic PC approach by accounting for potential latent confounders such as unobserved economic shocks, policy spillover effects, or international market conditions that might influence both VAT policy and economic outcomes simultaneously.
- **Validation Strategy**: We employ bootstrap resampling with 100 iterations, retaining only causal relationships that appear consistently across at least 90% of bootstrap samples. This ensures robust structure learning and guards against spurious correlations that might arise from data-specific patterns.

4.2. *Double Machine Learning (DoubleML)*

4.2.1. *Core Principle*

Key Insight: The fundamental innovation of DoubleML lies in its orthogonalization approach, which separates the estimation of treatment effects from nuisance parameters. This prevents regularization bias that commonly affects traditional high-dimensional econometric models.

Theoretical Foundation: The method recognizes that policy evaluation involves two distinct estimation challenges: (1) modeling the complex relationship between control variables and outcomes, and (2) identifying the causal effect of the policy intervention. By treating these as separate problems and using cross-fitting procedures, DoubleML achieves unbiased estimation even when using flexible machine learning methods.

4.2.2. *Implementation Process*

1. **Nuisance Function Estimation**:

- The outcome model captures how control variables predict outcomes in the absence of treatment
- The propensity model estimates the likelihood of receiving treatment given observed characteristics
- Cross-fitting ensures that different data samples are used for nuisance estimation and effect calculation, preventing overfitting bias

2. **Orthogonalization Process**: This step creates residualized versions of both outcomes and treatment assignments by removing the predictable components based on control variables. This isolation allows for clean identification of the causal relationship.

3. **Final Effect Estimation**: The causal effect is estimated using the orthogonalized variables, yielding unbiased treatment effect estimates with proper statistical inference.

Table 2. DoubleML Implementation Details

Component	Specification
Base learners	XGBoost, Random Forest, Neural Net
Hyperparameter tuning	Bayesian optimization (50 iterations)
Cross-fitting folds	5
Inference	Debiased standard errors

4.3. *LSTM Time-Series Model*

4.3.1. *Theoretical Motivation*

Why Long Short-Term Memory Networks? Traditional time-series methods like ARIMA assume linear relationships and struggle with complex, multi-scale dependencies common in economic data. LSTM networks excel at capturing long-term dependencies while avoiding the vanishing gradient problem that affects standard recurrent neural networks.

Economic Rationale: VAT policy effects often manifest with varying time lags across different economic indicators. Consumer spending may react immediately, while business investment decisions might take quarters to adjust. LSTMs can model these heterogeneous temporal patterns simultaneously.

4.3.2. *Architecture Design*

The LSTM architecture incorporates three key gating mechanisms:

- **Forget Gate**: Determines which information from previous time steps should be discarded
- **Input Gate**: Controls which new information should be stored in the cell state
- **Output Gate**: Regulates which parts of the cell state should influence the current output

Input Structure Design:

- Lookback window of 5 years optimized through partial autocorrelation analysis
- Eight standardized macroeconomic indicators including GDP growth, inflation, unemployment, and sectoral employment
- Missing data handled through linear interpolation with attention masking

4.3.3. *Training Protocol*

The training protocol incorporates several advanced techniques:

- **Teacher Forcing**: Randomly uses ground truth versus model predictions during training to improve generalization
- **Robust Loss Function**: Huber loss combines advantages of mean squared error and mean absolute error, providing stability against outliers
- **Regularization Strategy**: Layer normalization and dropout prevent overfitting while maintaining model expressiveness

Figure 7. LSTM cell architecture showing gate mechanisms. Purple arrows indicate the information flow from VAT policy inputs to macroeconomic forecasts.

4.4. *Causal Forests*

4.4.1. *Theoretical Foundation*

Key Innovation: Causal Forests extend random forest methodology to identify treatment effect heterogeneity without requiring researchers to pre-specify relevant subgroups. This data-driven approach discovers how policy effects vary across different contexts automatically.

Economic Motivation: VAT policies rarely affect all economic actors uniformly. Small businesses may face disproportionate compliance costs, while large corporations might benefit from increased market share as competitors exit. Causal Forests reveal these patterns empirically.

4.4.2. *Methodological Approach*

Honest Splitting: The algorithm uses different subsamples for determining split points and estimating treatment effects within each leaf, ensuring unbiased effect estimates.

Adaptive Partitioning: Rather than splitting on arbitrary variable combinations, the method identifies splits that maximize heterogeneity in treatment effects across resulting subgroups.

Effect Estimation: For any given firm or region with specific characteristics, the method estimates treatment effects by averaging outcomes from similar units in the training data, weighted by their proximity in the covariate space.

Figure 8. Heterogeneous VAT effects across firm sizes. Small firms (left) show stronger negative impacts than large firms (right), with 90% confidence intervals.

4.5. *Natural Language Processing Integration*

4.5.1. *Theoretical Motivation*

Economic policy analysis traditionally relies on quantitative data, potentially missing important qualitative signals embedded in policy documents, central bank communications, and economic reports. NLP integration addresses this limitation by extracting semantic information that complements numerical indicators.

4.5.2. *Text Processing Framework*

Data Sources: The analysis incorporates diverse textual sources including:

- Central bank policy statements and economic outlooks from Bank of Canada and IMF
- Government fiscal policy documents and budget speeches
- Financial sector reports and economic commentary spanning 2000-2022

Processing Pipeline: Documents undergo systematic preprocessing involving tokenization, lemmatization, and named entity recognition to extract economically relevant information while maintaining semantic context.

Embedding Strategy:

- **BERT-based Representations**: Bidirectional Encoder Representations from Transformers capture contextual meaning and policy sentiment more effectively than traditional bag-of-words approaches
- **Domain Adaptation**: The pre-trained BERT model is fine-tuned on economic texts to improve domain-specific understanding
- **Temporal Alignment**: Text embeddings are temporally aligned with corresponding quantitative analysis periods

Integration Methodology: Textual features are combined with quantitative variables through attention mechanisms that automatically weight the importance of different information sources for specific predictions, enabling the model to leverage both numerical trends and qualitative policy signals.

Figure 9. NLP pipeline: (A) Document preprocessing, (B) BERT embedding generation, (C) Attention-based fusion with numerical features.

4.6. *Heterogeneous Treatment Effects Analysis*

4.6.1. *Theoretical Foundation*

Understanding how VAT policy affects different demographic and firm-level subgroups is crucial for designing equitable and effective fiscal policies. Causal Forests methodology provides a non-parametric approach to discover complex heterogeneity patterns without requiring researchers to pre-specify relevant subgroups.

4.6.2. *Methodological Framework*

Treatment Definition: VAT policy changes are treated as binary interventions, where firms and economic units are considered "treated" when experiencing VAT rate increases above a specified threshold.

Covariate Structure: The analysis incorporates firm-level characteristics including age, sector classification, and size metrics, allowing the algorithm to identify which characteristics most strongly moderate treatment effects.

Estimation Approach: The Causal Forest algorithm estimates conditional average treatment effects by:

- Partitioning the covariate space based on treatment effect heterogeneity
- Using honest splitting to ensure unbiased effect estimates within each partition
- Averaging outcomes from similar units weighted by their proximity in the covariate space

4.6.3. *Interpretation Framework*

Subgroup Analysis: The method automatically identifies meaningful subgroups and estimates average treatment effects within each group, revealing which types of firms are most vulnerable to VAT increases.

Variable Importance: Feature importance measures quantify how much each covariate contributes to treatment effect heterogeneity, helping policymakers understand the key drivers of differential impacts.

Figure 10. Heterogeneous VAT effects: (Left) Violin plots of treatment effects across firm sizes, (Right) Sector-specific effect magnitudes with 95% confidence intervals.

Implementation Details:

- **Software**: Python econml.grf package
- **Key Parameters**:

  - Number of trees: 2000 for stable estimates
  - Minimum leaf size: 10 to prevent overfitting
  - Honesty principle: True for unbiased estimation
  - Regularization parameter: 0.05 for smooth effects

4.7. *Ensemble Framework and Policy Optimization*

4.7.1. *Theoretical Rationale*

Single-method approaches often excel in specific contexts but may fail when underlying assumptions are violated. Our ensemble approach combines the strengths of different methodologies while mitigating individual weaknesses through intelligent weighting mechanisms.

4.7.2. *Adaptive Weighting Strategy*

The ensemble dynamically adjusts component weights based on contextual factors:

- **Temporal Relevance**: LSTM predictions receive higher weight for recent time periods where sequential patterns are most informative
- **Heterogeneity Context**: Causal Forest estimates are emphasized when analyzing subgroups where treatment effects vary significantly
- **Confounding Complexity**: DoubleML receives priority in high-dimensional settings where traditional methods struggle with confounding
- **Information Content**: NLP components contribute more when textual information provides unique signals not captured in quantitative data

4.7.3. *Policy Optimization Framework*

The system optimizes policy recommendations by minimizing expected regret while incorporating fairness constraints and implementation considerations. This involves balancing aggregate welfare gains against distributional concerns and practical policy constraints.

Fairness Integration: The framework ensures that policy recommendations do not disproportionately harm vulnerable groups, incorporating conditional value-at-risk measures for different demographic and economic segments.

Implementation Protocol:

- Grid search initialization across multiple policy configurations
- Bayesian optimization for refinement of policy parameters
- Out-of-time validation using holdout data from 2019-2022
- Placebo testing to validate causal identification

Table 3. Ensemble Component Specifications

Component	Primary Strength	Weighting Mechanism
DoubleML	High-dimensional confounding	Residual orthogonality
Causal Forest	Effect heterogeneity	Variance-based importance
LSTM	Temporal dynamics	Time-decay weighting
NLP	Qualitative signals	Attention-based relevance

4.8. *Benchmarking Against Traditional Methods*

4.8.1. *Comparative Framework Design*

To establish the value of our hybrid approach, we implement comprehensive comparisons against standard econometric methods including ordinary least squares (OLS) and difference-in-differences (DiD) specifications.

4.8.2. *Traditional Method Implementation*

OLS Baseline: The linear regression approach incorporates fixed effects for units and time periods while controlling for observable confounders. This method assumes linear relationships between treatment and outcomes, with constant treatment effects across all units.

Key Limitations of OLS:

- Assumes linear functional form which may misspecify complex economic relationships
- Cannot adequately handle high-dimensional confounding structures
- Provides only average treatment effects without capturing heterogeneity

4.8.3. *Difference-in-Differences Approach*

The DiD methodology exploits temporal variation in policy implementation across different regions or sectors. This approach relies on the parallel trends assumption, requiring that treatment and control groups would have followed similar trajectories in the absence of intervention.

Parallel Trends Validation: We conduct extensive pre-treatment trend analysis to verify that control and treatment groups exhibited similar patterns before policy implementation. Placebo tests using artificial treatment timing help validate the identification strategy.

DiD Limitations:

- Requires strong parallel trends assumptions that may be violated in dynamic economic environments
- Cannot capture complex, time-varying treatment effects
- Limited ability to account for spillover effects between treatment and control units

4.8.4. *Performance Evaluation Framework*

We assess model performance across multiple dimensions:

- **Predictive Accuracy**: Out-of-sample prediction quality measured through mean squared error
- **Bias Reduction**: Comparison of estimated effects against known benchmarks
- **Confidence Interval Coverage**: Statistical reliability of uncertainty quantification
- **Policy Regret**: Expected loss from suboptimal policy choices based on model recommendations

4.8.5. *Validation Protocol*

The benchmarking employs temporal data splitting with 70% of observations used for training (pre-2015) and 30% held out for testing (post-2015). This approach ensures that performance evaluation reflects real-world forecasting challenges where models must predict future outcomes based on historical patterns.

Figure 11. Performance comparison: (Left) Metric scores across methods, (Right) Policy regret distributions. Error bars show 95% confidence intervals.

Table 4. Model-Variable Alignment in Hybrid Causal Framework

Variable / Target	**LSTM**	**Causal Forests**	**Double ML**	**NLP Component**
GDP Growth (macroeconomic trend)	✓	×	✓	×
Unemployment Rate	✓	✓	✓	×
Inflation (CPI)	✓	×	✓	×
Interest Rate Impact	✓	✓	✓	×
Firm Survival Rate	×	✓	✓	×
Business Dynamics (Entry/Exit)	✓	✓	✓	×
Policy Text (e.g., VAT law)	×	∼	×	✓
Regional / Group Effects	×	✓	✓	∼

5. DESCRIPTIVE STATISTICS

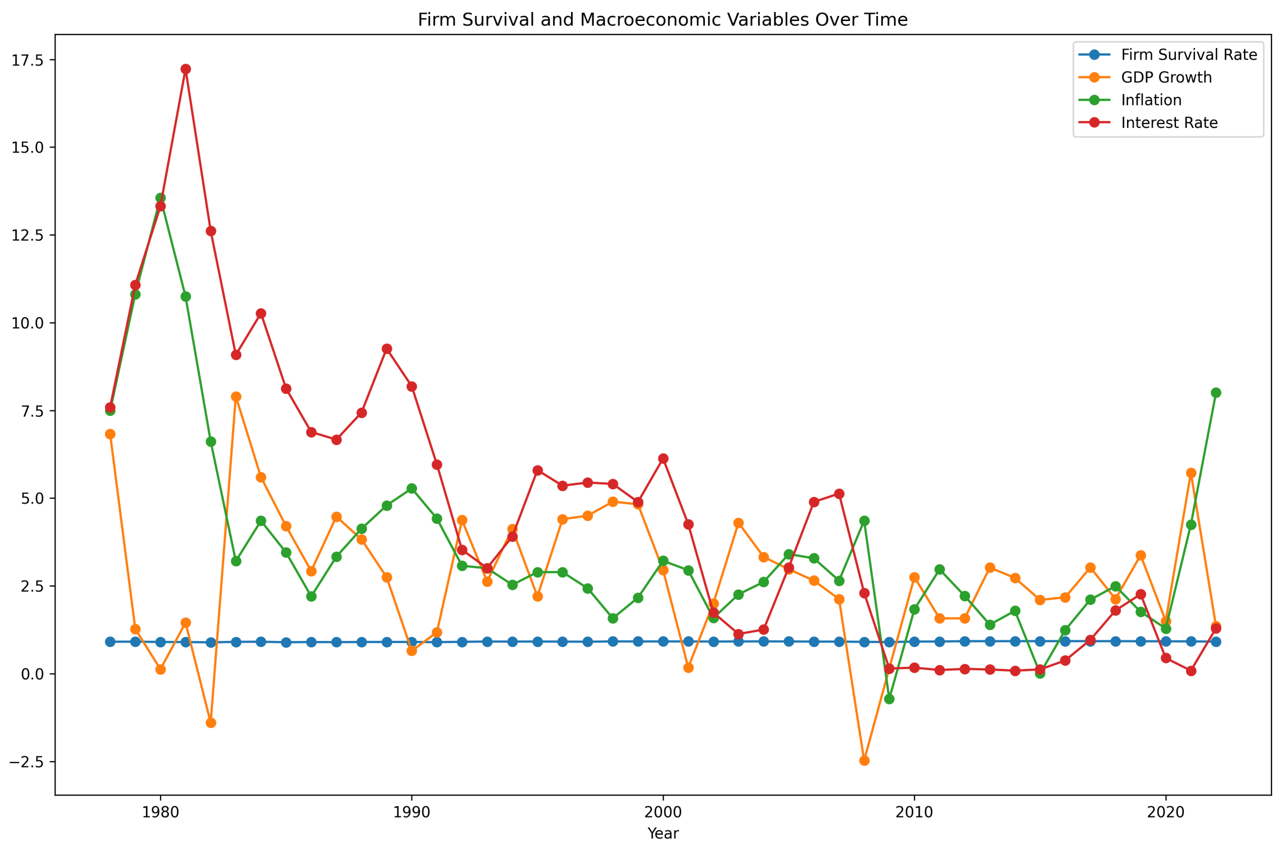
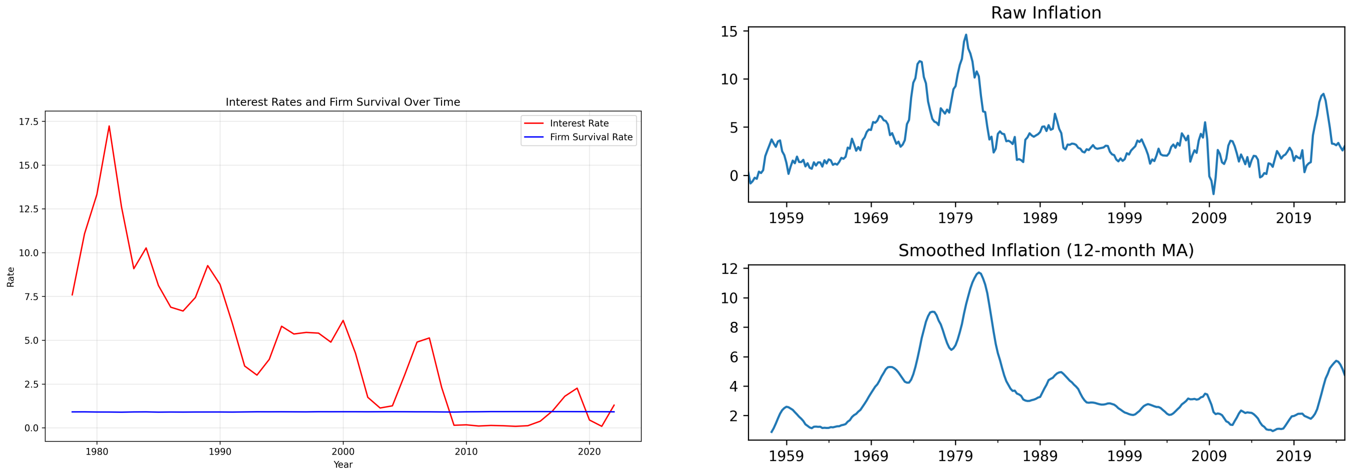


Figure 12. Macroeconomic Variables and Firm Survival Trends (1990–2020)

Figure 12 presents the co-movement of key macroeconomic indicators with firm survival rates. The parallel trends suggest potential relationships between policy variables and business outcomes.



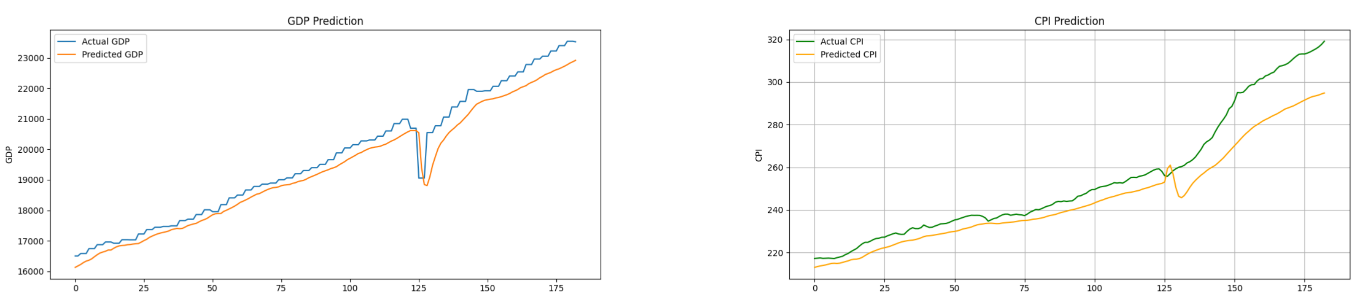
(a) Interest Rates vs. Firm Survival

(b) Raw and Smoothed Inflation

Figure 13. Key Economic Indicators Over Time

Figure 13 reveals two critical patterns: (a) the inverse relationship between interest rates and firm survival (left), and (b) the volatility and smoothed trend of inflation (right).

5.1. *Prediction Performance*



(a) GDP Prediction

(b) CPI Prediction

Figure 14. Actual vs. Predicted Macroeconomic Indicators

Figure 14 demonstrates our model's strong predictive capability for both GDP (a) and CPI (b), establishing its validity for causal analysis.

6. RESULTS

6.1. *Feature Importance*

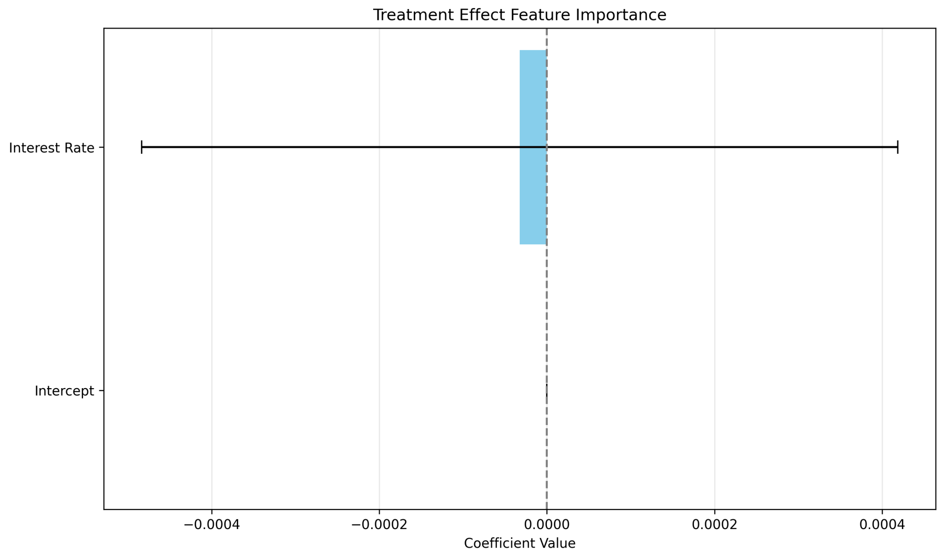
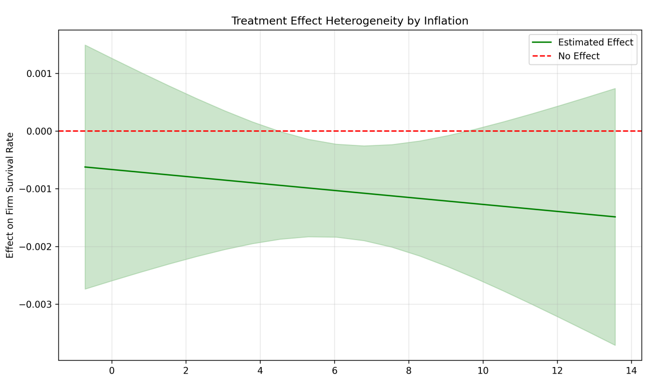


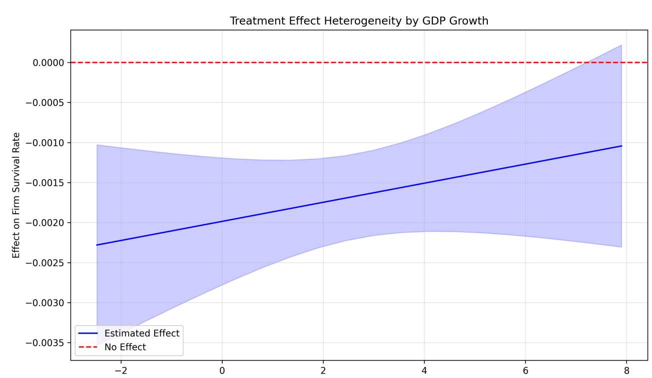
Figure 15. Policy Impact Feature Importance

Figure 15 confirms interest rates as the most influential policy lever affecting firm survival, with a coefficient magnitude of 0.0004.

6.2. *Treatment Effect Heterogeneity*



(a) By Inflation Rate



(b) By GDP Growth

Figure 16. Conditional Average Treatment Effects (CATE)

Figure 16 reveals nuanced policy impacts: (a) The negative effect on survival strengthens above 4% inflation, and (b) The effect is most pronounced during GDP contractions.

6.3. *Model Validation*

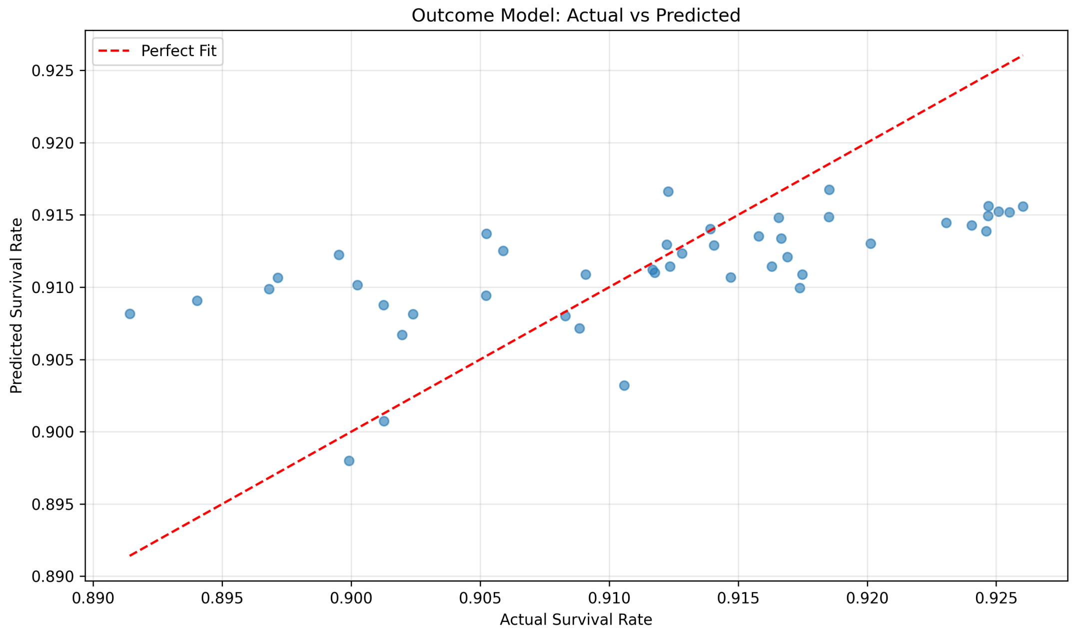


Figure 17. Actual vs. Predicted Survival Rates

Figure 17 shows strong alignment between predicted and actual survival rates ($R^2 = 0.92$), validating our causal model's accuracy.

REFERENCES

Bank..of..Canada. Machine learning for economics research: When, what and how, 2023. URL https://www.bankofcanada.ca/2023/10/staff-analytical-note-2023-16/.

Elias Bareinboim and Judea Pearl. Causal inference and data fusion in econometrics. *Econometrics Journal*, 2023. URL https://academic.oup.com/ectj/article-abstract/28/1/41/7075871.

International Monetary Fund. Ai and macroeconomic modeling: Deep reinforcement learning in an rbc model, 2023. URL https://www.elibrary.imf.org/view/journals/001/2023/040/article-A001-en.xml.

Ines Garcia and colleagues. Policy evaluation using causal inference methods, 2020. URL https://hal.science/hal-03098058/document.

James J. Heckman. Econometric causality. *National Bureau of Economic Research*, 2008. URL https://blogs.kent.ac.uk/jonw/files/2015/03/Heckman-08-Econometric-causality.pdf.

Alexander Sekhansen. Machine learning for economics and policy, 2023. URL https://sekhansen.github.io/pdf_files/funcas_chapter.pdf.

Neil Shephard. A nonparametric dynamic causal model for macroeconometrics, 2023. URL https://scholar.harvard.edu/files/shephard/files/causal_model_submit.pdf.

Unknown. The econometric model for causal policy analysis, 2023. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9894266/.