

**UNIVERSITÉ NATIONALE DU VIETNAM, HANOÏ
ÉCOLE INTERNATIONALE**

=====



BIABA KUYA JIRINCE

**ANALYSE CAUSALE DE L'IMPACT
MACROÉCONOMIQUE À PARTIR
DE SÉRIES TEMPORELLES À L'AIDE
DU CAUSAL MACHINE LEARNING**

**PHÂN TÍCH NHÂN QUẢ VỀ TÁC ĐỘNG
KINH TẾ VĨ MÔ TỪ CHUỖI THỜI GIAN
BẰNG HỌC MÁY NHÂN QUẢ**

Programme : Systèmes Intelligents et Multimédia

Code : 8480201.02

Encadrant(s) : Prof. Selain KASEREKA KABUNGA
Dr. Đào Tùng

MASTER SYSTÈMES INTELLIGENTS ET MULTIMÉDIA

Kinshasa – 2026

**UNIVERSITÉ NATIONALE DU VIETNAM, HANOÏ
ÉCOLE INTERNATIONALE**

=====



BIABA KUYA JIRINCE

**ANALYSE CAUSALE DE L'IMPACT
MACROÉCONOMIQUE À PARTIR
DE SÉRIES TEMPORELLES À L'AIDE
DU CAUSAL MACHINE LEARNING**

**PHÂN TÍCH NHÂN QUẢ VỀ TÁC ĐỘNG
KINH TẾ VĨ MÔ TỪ CHUỖI THỜI GIAN
BẰNG HỌC MÁY NHÂN QUẢ**

Programme : Systèmes Intelligents et Multimédia

Code : 8480201.02

Encadrant(s) : Prof. Selain KASEREKA KABUNGA
Dr. Đào Tùng

MASTER SYSTÈMES INTELLIGENTS ET MULTIMÉDIA

Kinshasa – 2026

Attestation sur l'honneur

★

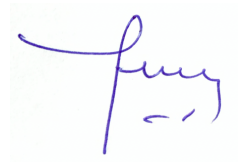
Je déclare par la présente que ce mémoire a été réalisé par moi-même sous la direction et la supervision du Professeur Selain KASEREKA KABUNGA, du Professeur Christian MULOMBA et du Professeur Nathanael KASORO; et que les travaux et les résultats qu'il contient sont conformes à la réalité de l'auteur et respectent l'éthique de la recherche. Les données et figures présentées dans ce mémoire sont issues de mes propres analyses, commentaires et évaluations, issues de diverses sources et sont dûment mentionnées dans la partie référence.

De plus, les autres commentaires, analyses et données utilisés par d'autres auteurs et organisations ont été explicitement mentionnés.

J'assume l'entière responsabilité de toute fraude détectée dans mon mémoire.

Kinshasa, 08 Juin 2026

(Auteur)



BIABA KUYA JIRINCE

Approuvé par

ENCADRANT



(Signature et nom complet)

Prof. Selain KASEREKA KABUNGA

ENCADRANT



(Signature et nom complet)

Dr. Đào Tùng

Épigraphe

“The task of causal inference is to extract information about cause-and-effect relationships from data.”

Judea Pearl

Remerciements

Je remercie tout d'abord le Créateur de l'univers, Dieu Tout-Puissant, pour la grâce, la sagesse et la force accordées tout au long de cette période de formation et de recherche.

J'exprime ma profonde gratitude à mon directeur de stage, **Professeur Selain KASEREKA KABUNGA**, pour sa disponibilité, ses orientations scientifiques rigoureuses et son engagement constant dans le suivi de ce travail.

Je remercie sincèrement mon co-encadrant, **Professeur Christian MULOMBA**, dont les contributions précieuses à l'orientation méthodologique, les conseils avisés sur le Causal Machine Learning et le Double Machine Learning, et l'expertise en analyse macroéconomique ont été déterminants dans la réussite de cette recherche.

J'adresse ma reconnaissance au corps enseignant de l'Institut Francophone International (IFI-VNU), dont l'excellence académique et la rigueur scientifique ont constitué le socle de ma formation en Systèmes Intelligents et Multimédia. Je remercie en particulier **Dr. Vinh HO TUONG** pour son accompagnement pédagogique tout au long de ma formation et ses conseils précieux sur la rédaction scientifique.

Je tiens à exprimer mon entière gratitude au Professeur **Nathanael KASORO MULENDA**, Directeur du laboratoire ABIL et co-encadrant de ce travail. Je remercie également l'ensemble des membres du laboratoire ABIL de l'Université de Kinshasa pour l'environnement de recherche stimulant et collaboratif mis à ma disposition durant ce stage.

Je tiens à exprimer ma profonde reconnaissance à mon père, **Lambert BIABA KUYA**, à ma mère, **Marthe KIANA NGUMBANDA**, ainsi qu'à ma sœur, **Love BIABA NGALEY**, pour leur soutien inconditionnel, leur patience et leurs sacrifices, qui ont constitué le pilier de mon parcours académique.

Je remercie également du fond du cœur **Isabelle BILEIBA POSHO** pour son soutien indéfectible, sa patience et ses encouragements quotidiens tout au long de mon parcours de Master au Vietnam.

Jirince BIABA KUYA

Dédicace

*À Marthe Désirée Olivia,
pour son soutien indéfectible durant cette
période exigeante, je dédie ce travail.*

Jirince BIABA KUYA

Résumé

Ce travail de stage porte sur l'analyse causale de l'impact macroéconomique de la Formation Brute de Capital sur la croissance du PIB à partir de séries temporelles annuelles, en utilisant des techniques avancées de Causal Machine Learning (CML). L'étude couvre un panel hétérogène de 7 pays, notamment la France, l'Allemagne, le Maroc, le Ghana, le Nigeria, l'Angola et la République Démocratique du Congo sur la période 2000–2024 (175 observations), avec des données harmonisées issues du Fonds Monétaire International (FMI WEO, FMI IFS) et de la Banque Mondiale.

La méthodologie s'articule en deux phases complémentaires. La Phase 1 applique quatre méthodes de Causal Discovery (test de causalité de Granger, modèle VAR avec IRF et FEVD, modèle VECM avec test de Johansen, et algorithme PCMCI avec test CMlknn non linéaire pour identifier les structures causales entre variables macroéconomiques). Les résultats de ces méthodes sont synthétisés en un graphe causal (DAG) de consensus multi-méthodes avec score de support pour chaque pays. La Phase 2 applique le Double Machine Learning (LinearDML et CausalForestDML, bibliothèque econml) pour estimer l'effet causal quantifié de la Formation de Capital sur la croissance du PIB, en contrôlant rigoureusement cinq confondeurs macroéconomiques (inflation, dette publique, balance commerciale, taux de change, réserves de change) et les effets fixes pays.

Les résultats du Causal Discovery révèlent que le lien Formation de Capital $\rightarrow$ PIB est absent ou fragile dans les 7 pays, le taux de change s'imposant comme la variable causale dominante dans 5 pays sur 7.

Le Double Machine Learning estime un ATE global de -0.528 (IC 95% : $[-1.042; -0.014]$, $p = 0.044$), négatif et significatif, masquant deux régimes causaux opposés : un effet négatif significatif en France (ATE = -3.728 , $p = 0.023$) cohérent avec les rendements décroissants du capital dans les économies matures, et un effet positif significatif en RDC (ATE = $+0.542$, $p = 0.022$) cohérent avec le déficit structurel de capital dans les économies à faible revenu. Le test de Welch confirme une hétérogénéité structurelle hautement significative entre pays développés et pays en développement ($t = -6.965$, $p = 4.17 \times 10^{-9}$).

Ces résultats démontrent que l'effet de l'investissement sur la croissance n'est pas universel et dépend fondamentalement du niveau de développement et du stock de capital existant. À notre connaissance, ce travail constitue l'une des premières applications empiriques combinant Causal Discovery et Double Machine Learning sur un panel macroéconomique couvrant simultanément des économies européennes et africaines, et pose les bases d'un second article scientifique empirique.

Mots-clés : Causal Machine Learning, Double Machine Learning, Causal Discovery, Séries temporelles macroéconomiques, Formation de Capital, Croissance économique, Hétérogénéité, Afrique subsaharienne, PCMCI, DAG.

Abstract

This work focuses on the causal analysis of the macroeconomic impact of Gross Capital Formation on GDP growth using annual time series data and advanced Causal Machine Learning (CML) techniques. The study covers a heterogeneous panel of 7 countries, France, Germany, Morocco, Ghana, Nigeria, Angola, and the Democratic Republic of Congo over the period 2000–2024 (175 observations), using harmonized data from the International Monetary Fund (IMF WEO, IMF IFS) and the World Bank.

The methodology is structured in two complementary phases. Phase 1 applies four Causal Discovery methods Granger causality test, VAR model with IRF and FEVD, VECM model with Johansen cointegration test, and PCMCI algorithm with nonlinear CMiknn test to identify causal structures among macroeconomic variables. The results of these methods are synthesized into a consensus multi-method causal graph (DAG) with a support score for each country. Phase 2 applies Double Machine Learning (LinearDML and CausalForestDML, econml library) to estimate the quantified causal effect of Capital Formation on GDP growth, rigorously controlling for five macroeconomic confounders (inflation, government debt, trade balance, exchange rate, foreign exchange reserves) and country fixed effects.

The Causal Discovery results reveal that the Capital Formation $\rightarrow$ GDP link is absent or fragile across all 7 countries, with the exchange rate emerging as the dominant causal variable in 5 out of 7 countries. The Double Machine Learning estimates a global ATE of -0.528 (95% CI : $[-1.042; -0.014]$, $p = 0.044$), negative and significant, concealing two structurally opposed causal regimes : a significant negative effect in France (ATE = -3.728 , $p = 0.023$) consistent with diminishing returns to capital in mature economies, and a significant positive effect in the DRC (ATE = $+0.542$, $p = 0.022$) consistent with structural capital deficits in low-income economies. The Welch test confirms a highly significant structural heterogeneity between developed and developing countries ($t = -6.965$, $p = 4.17 \times 10^{-9}$).

These results demonstrate that the effect of investment on growth is not universal and depends fundamentally on the level of development and the existing capital stock. To our knowledge, this work constitutes one of the first empirical applications combining Causal Discovery and Double Machine Learning on a macroeconomic panel simultaneously covering European and African economies, and lays the groundwork for a second empirical scientific article.

Keywords : Causal Machine Learning, Double Machine Learning, Causal Discovery, Macroeconomic time series, Capital Formation, Economic growth, Heterogeneity, Sub-Saharan Africa, PCMCI, DAG.

Liste des Acronymes

AIC	<i>Akaike Information Criterion</i>
ATE	<i>Average Treatment Effect</i>
BIC	<i>Bayesian Information Criterion</i>
CF	<i>Capital Formation</i>
CML	<i>Causal Machine Learning</i>
CMiknn	<i>Conditional Mutual Information k-Nearest Neighbors</i>
CV	<i>Cross-Validation</i>
DAG	<i>Directed Acyclic Graph</i>
DML	<i>Double Machine Learning</i>
DSGE	<i>Dynamic Stochastic General Equilibrium</i>
FBC	<i>Formation Brute de Capital</i>
FDR	<i>False Discovery Rate</i>
FEVD	<i>Forecast Error Variance Decomposition</i>
FMI	<i>Fonds Monétaire International</i>
FPE	<i>Final Prediction Error</i>
GDP	<i>Gross Domestic Product</i>
HQIC	<i>Hannan-Quinn Information Criterion</i>
HTE	<i>Heterogeneous Treatment Effects</i>
IC	<i>Intervalle de Confiance</i>
IFS	<i>International Financial Statistics</i>
IRF	<i>Impulse Response Function</i>
ML	<i>Machine Learning</i>
PCMCI	<i>Peter and Clark Momentary Conditional Independence</i>
PIB	<i>Produit Intérieur Brut</i>
RDC	<i>République Démocratique du Congo</i>
SUTVA	<i>Stable Unit Treatment Value Assumption</i>
VAR	<i>Vector Autoregression</i>
VECM	<i>Vector Error Correction Model</i>
WEO	<i>World Economic Outlook</i>

Table des matières

Attestation sur l'honneur

Remerciements

Dédicace i

Résumé ii

Abstract iii

Liste des Acronymes iv

Table des matières v

Liste des tableaux ix

Liste des figures 1

Publications 2

Introduction Générale 3

1 Contexte et motivation 3

2 Problématique 4

3 Questions de recherche 4

4 Hypothèses de recherche 5

5 Objectifs 5

6 Démarche méthodologique 6

7 Contribution de l'étude 6

8 Cadre institutionnel d'accueil 7

9 Structure du mémoire 7

1 Cadre Théorique et État de l'Art 9

1.1 Introduction..... 9

1.2 Inférence Causale 9

1.2.1 Le Framework des Résultats Potentiels 9

1.2.2 Les Modèles Causaux Graphiques (DAGs)..... 9

1.3 Causalité de Granger et Extensions 10

1.3.1 Test de Causalité de Granger 10

1.3.2 Limites du Test de Granger 10

1.4 Modèle VAR et Analyse des Dynamiques..... 10

1.4.1	Définition et Formulation	10
1.4.2	Fonction de Réponse Impulsionnelle (IRF)	11
1.4.3	Décomposition de la Variance (FEVD)	11
1.5	Cointégration et Modèle VECM	11
1.5.1	Test de Cointégration de Johansen	11
1.5.2	Modèle VECM	11
1.6	Découverte Causale Non Linéaire : PCMCI	12
1.6.1	Principe et Motivation	12
1.6.2	Formulation	12
1.6.3	Test CMlkn	12
1.7	Double Machine Learning	13
1.7.1	Motivation et Principe	13
1.7.2	Formulation du Modèle Partiellement Linéaire	13
1.7.3	Cross-Fitting	13
1.7.4	Causal Forest et Effets Hétérogènes	14
1.7.5	Hypothèses d'Identification	14
1.7.6	Spécification du Modèle dans Notre Travail	14
1.8	Limites des Modèles Économétriques Classiques	15
1.9	Émergence du Causal Machine Learning en Macroéconomie	15
1.9.1	Du Prédictif au Causal	15
1.9.2	Applications en Macroéconomie	15
1.9.3	Adoption Fragmentée et Gap Africain	16
1.10	Outils, Librairies et Algorithmes du CML	16
1.11	Applications en Contexte Africain	17
1.12	Travaux sur l'Effet de l'Investissement sur la Croissance	18
1.12.1	Perspective Théorique	18
1.12.2	Études Empiriques sur Données de Panel Africain	18
1.13	Positionnement de Notre Travail dans l'État de l'Art	19
1.13.1	Ce que la Littérature a Établi	19
1.13.2	Le Gap que Notre Travail Comble	19
1.13.3	Notre Contribution Spécifique	19
1.14	Conclusion partielle	19
2	Données et Méthodologie	21
2.1	Introduction	21
2.2	Sources de Données	21
2.2.1	Panel de Pays et Période	21
2.2.2	Sources Primaires	21
2.2.3	Les 7 Variables Macroéconomiques	21
2.2.4	Architecture du système	22
2.3	Prétraitement des Données	23
2.3.1	Harmonisation des Sources	23
2.3.2	Transformations Statistiques	24

2.3.3	Winsorisation	24
2.3.4	Standardisation Z-score	24
2.4	Structure Temporelle Causale	25
2.4.1	Justification du Lagging	25
2.4.2	Fichiers Produits	25
2.4.3	Impact sur les Observations	25
2.5	Phase 1 : Pipeline de Causal Discovery	26
2.5.1	Vue d'Ensemble	26
2.5.2	Étape 1 : Test de Causalité de Granger	26
2.5.3	Étape 2 : Modèle VAR	26
2.5.4	Étape 3 : Test de Johansen et VECM	26
2.5.5	Étape 4 : PCMCI	26
2.5.6	Étape 5 : Synthèse DAG Multi-Méthodes	27
2.6	Phase 2 : Pipeline Double Machine Learning	27
2.6.1	Vue d'Ensemble	27
2.6.2	Construction de la Matrice DML	27
2.6.3	Évaluation des Modèles Nuisances	28
2.6.4	Estimation de l'ATE Global	29
2.6.5	Estimation par Pays	29
2.6.6	Estimation par Groupe	29
2.6.7	Hétérogénéité Individuelle : CausalForestDML	30
2.7	Conclusion partielle	30
3	Résultats et Discussion	31
3.1	Introduction	31
3.2	Analyse Descriptive des Données	31
3.2.1	Évolution Temporelle des Variables	31
3.2.2	Matrice de Corrélation	33
3.2.3	Relation Capital Formation et GDP Growth	34
3.3	Résultats de la Phase 1 : Causal Discovery	34
3.3.1	Résultats du Test de Granger	34
3.3.2	Résultats du Modèle VAR	36
3.3.3	Résultats du Test de Johansen et du VECM	37
3.3.4	Résultats du PCMCI	38
3.3.5	Synthèse DAG Multi-Méthodes	40
3.4	Résultats de la Phase 2 : Double Machine Learning	45
3.4.1	ATE Global	45
3.4.2	ATE par Pays	45
3.4.3	ATE par Groupe de Développement	47
3.4.4	Effets Hétérogènes Individuels : CausalForestDML	48
3.4.5	Discussion Générale des Résultats DML	49
3.4.6	Discussion de la robustesse des estimations	50
3.5	Conclusion partielle	51

Liste des tableaux

1	Bureau du Conseil de Gestion d'ABIL (Source ABIL)	7
1.1	Principaux outils et méthodes du CML pour les séries temporelles macroéconomiques.	17
1.2	Comparaison des principaux algorithmes de découverte causale pour séries temporelles	17
2.1	Description des variables macroéconomiques utilisées	22
2.2	Transformations appliquées par variable	24
2.3	Statistiques descriptives globales (données standardisées, $n = 174^1$)	24
2.4	Performance des modèles nuisances en validation croisée 5-fold	28
3.1	Résultats du test de Granger par pays (lag 1–3, $\alpha = 0.05$)	35
3.2	Lag optimal sélectionné par critère BIC pour chaque pays	36
3.3	Test de Johansen : statistiques de trace et valeurs critiques à 95% — France (rang plein $r = k - 1 = 6$, avec $k = 7$ variables)	38
3.4	Résultats PCMCi par pays ($\tau_{max} = 3$, $\alpha = 0.05$, CMiknn, FDR BH)	39
3.5	Synthèse des DAGs finaux pour les 7 pays (Causal Discovery multi-méthodes).....	44
3.6	ATE global avec LinearDML (167 observations, 7 pays, 2000-2024)	45
3.7	ATE par pays avec LinearDML	45
3.8	ATE par groupe (LinearDML avec BootstrapInference).....	47
3.9	HTE par groupe avec CausalForestDML (800 arbres)	48
3.10	Test de Welch : Différence des HTE entre groupes.....	48
3.11	Comparaison LinearDML vs CausalForestDML par groupe	49

Liste des figures

2.1	Architecture séquentielle de la méthodologie de recherche combinant la découverte causale et le double apprentissage automatique pour l'analyse de la croissance du PIB dans sept pays (2000–2024).	23
2.2	Comparaison des performances des modèles nuisances en validation croisée 5-fold (R^2 pour les équations outcome $E[Y X]$ et traitement $E[T X]$). Le modèle retenu est encadré en bleu.	29
3.1	Évolution du taux de croissance annuel du PIB réel par pays (2000–2024, valeurs standardisées, Z-score). Source : FMI WEO et Banque Mondiale.	31
3.2	Évolution de la Formation Brute de Capital par pays (2000–2024, valeurs standardisées, Z-score). Source : FMI WEO et Banque Mondiale.	32
3.3	Matrice de corrélation des 7 variables standardisées ($n = 174$).	33
3.4	Nuage de points Capital_Formation vs GDP_Growth (valeurs standardisées, Z-score).	34
3.5	Heatmap des p-values Granger minimales (lags 1–3) Allemagne.	35
3.6	Décomposition de la variance de l'erreur de prévision (FEVD) de l'Allemagne (lag=2, horizon=10).	37
3.7	Graphe causal temporel PCMCI : Nigeria (8 liens significatifs).	39
3.8	Graphe causal temporel PCMCI : Ghana.	40
3.9	DAG final de la France (support ≥ 1 , acyclique).	41
3.10	DAG final de l'Angola (support ≥ 1 , acyclique).	42
3.11	DAG final du Nigeria (support ≥ 1 , acyclique).	43
3.12	Effets causaux moyens (ATE) par pays, LinearDML avec IC 95%.	46
3.13	ATE par groupe de développement avec LinearDML (IC 95%).	47
3.14	Distribution des effets de traitement hétérogènes individuels $\tau(x_i)$ avec CausalForestDML.	49

Publications

Les travaux conduits dans le cadre de ce mémoire ont donné lieu à des publications et communications scientifiques dans des revues et conférences internationales à comité de lecture. Ces contributions témoignent de l’ancrage de la recherche dans la littérature académique et de sa diffusion auprès de la communauté scientifique spécialisée en intelligence artificielle appliquée à l’analyse macroéconomique.

1. **Biaba, J. K., Mulomba, C. M., Bazie, I. G., Nzazi, B. N., Tashev, T., Kambale, W. V., ... & Kasereka, S. K.** (2025). Causal Machine Learning Applied to Macroeconomic Analysis : Applications, Challenges, and Perspectives. *Procedia Computer Science*, **272**, 251–260.
<https://doi.org/10.1016/j.procs.2025.10.203>
2. **Biaba, J. K., Mulomba, C. M., Kasoro, N. M., & Kasereka, S. K.** (2026). Beyond Correlation : Quantifying the Causal Effect of Investment on Growth through Causal Discovery and Double Machine Learning. Soumis à *IEEE Access* (Q1). (*Sous révision.*)

Introduction Générale

Cette introduction est organisée en neuf sections. La Section 1 présente le contexte et la motivation de ce travail. La Section 2 formule la problématique centrale. La Section 3 énonce les questions de recherche. La Section 4 propose les hypothèses correspondantes. La Section 5 détaille les objectifs poursuivis. La Section 6 décrit la démarche méthodologique adoptée. La Section 7 présente les contributions originales de l'étude. La Section 8 décrit le cadre institutionnel d'accueil. La Section 9 expose la structure du mémoire.

1 Contexte et motivation

L'analyse macroéconomique est au cœur de la politique économique et de la planification du développement. Comprendre les mécanismes qui lient l'investissement, la croissance, l'inflation et la dette publique conditionne la qualité des décisions de politique publique, dans les économies avancées comme dans les pays en développement. Les modèles économétriques traditionnels, VAR (*Vector Autoregression*), VECM (*Vector Error Correction Models*) et DSGE (*Dynamic Stochastic General Equilibrium*), ont longtemps dominé cette analyse en imposant des hypothèses de linéarité, de stationnarité et d'exogénéité. Ces contraintes structurelles réduisent leur capacité à capturer la complexité des systèmes économiques réels [Shojaie and Fox \(2022\)](#); [Maitra \(2025\)](#). Loin d'être de simples artefacts techniques, elles conduisent à des biais d'identification dont les effets sur les recommandations de politique économique peuvent être substantiels.

Le *Causal Machine Learning* (CML) émerge comme une réponse directe à ces limites, combinant la puissance prédictive de l'apprentissage automatique avec la rigueur de l'inférence causale [Guo et al. \(2020\)](#). Fondé sur deux cadres théoriques complémentaires, le *framework des résultats potentiels* (Rubin, 2007) et les *modèles causaux graphiques* (Pearl), le CML permet d'estimer des effets causaux à partir de données observationnelles sans nécessiter de randomisation expérimentale. Cette propriété le rend particulièrement adapté au contexte macroéconomique, où les expériences contrôlées sont structurellement impossibles [Oprescu et al. \(2019\)](#).

Ce travail a été réalisé lors d'un stage de recherche au sein du laboratoire ABIL (*Artificial intelligence, BIg data and modeLing simulation*) de l'Université de Kinshasa (UNIKIN), République Démocratique du Congo, en collaboration avec l'École Internationale de l'Université Nationale du Vietnam, Hanoï (IFI-VNU). Il s'inscrit dans la continuité d'une revue systématique publiée à la conférence internationale EUSPN 2025 [Biaba et al. \(2025\)](#). Cette revue a identifié à la fois le potentiel et les lacunes de l'application du CML en macroéconomie, notamment en Afrique subsaharienne.

Le contexte économique mondial choisi présente une diversité structurelle d'une grande richesse analytique. La France et l'Allemagne font face à des défis de soutenabilité de la dette publique et de rendements décroissants de l'investissement. Les économies subsahariennes retenues, Angola, Ghana, Nigeria et République Démocratique du Congo (RDC), sont confrontées à des déficits structurels de capital, une forte volatilité macroéconomique et une dépendance aux matières premières. Le Maroc

représente un cas intermédiaire d'économie émergente en transition, enrichissant la comparaison sur trois régimes de développement distincts. Cette diversité géographique et structurelle constitue une opportunité scientifique rare pour tester si l'effet causal de l'investissement sur la croissance est universel ou structurellement différencié selon le niveau de développement.

2 Problématique

La relation entre la Formation Brute de Capital (FBC) et la croissance du produit intérieur brut (PIB) est l'une des plus étudiées en économie du développement. Depuis [Solow \(1956\)](#) pour la croissance exogène et [Romer \(1990\)](#) pour la croissance endogène, la théorie économique prédit un effet positif de l'investissement sur la croissance, à travers l'accumulation du capital physique et le progrès technologique. Cette prédiction, robuste dans le cadre théorique, ne se reflète pourtant pas systématiquement dans les données observationnelles.

Ce décalage est particulièrement marqué pour les pays africains. Des variables confondantes, inflation, taux de change et dette publique, influencent simultanément l'investissement et la croissance. Des effets non linéaires, absents des modèles classiques, opèrent par ailleurs entre ces variables. La causalité inverse pose un problème supplémentaire : la croissance peut stimuler l'investissement plutôt que l'inverse. L'hétérogénéité structurelle entre pays développés et pays en développement complique davantage toute généralisation.

Devant ces obstacles, les méthodes classiques d'économétrie, régression linéaire, test de Granger bivarié et modèle VAR, peinent à identifier l'effet causal *net* de l'investissement sur la croissance. Elles mesurent des corrélations, non des causes. Or une politique d'investissement efficace suppose de connaître la magnitude et la direction de l'effet causal, et non la seule association statistique.

De ce constat découle la question centrale de ce travail : les méthodes de Causal Machine Learning permettent-elles d'identifier et de quantifier l'effet causal net de la formation de capital sur la croissance du PIB, en contrôlant les confondeurs macroéconomiques ? Cet effet varie-t-il de manière structurelle selon le niveau de développement économique ? Deux sous-problèmes complémentaires la précisent. Le premier porte sur la structure causale : quelles variables macroéconomiques causent réellement la croissance du PIB dans chacun des 7 pays étudiés, et le capital joue-t-il un rôle direct ou indirect, médiatisé par d'autres variables comme le taux de change ou la dette publique ? Le second porte sur la quantification causale : quelle est la magnitude de l'effet causal net de la formation de capital sur la croissance, une fois contrôlés tous les confondeurs, et cet effet varie-t-il entre pays développés et pays en développement ?

3 Questions de recherche

À partir de la problématique formulée, trois questions de recherche structurent ce travail.

La première question est la suivante : les méthodes de Causal Discovery multi-méthodes (Granger, VAR, VECM, PCMCI) permettent-elles d'identifier des structures causales stables et interprétables entre variables macroéconomiques, dans des contextes économiques aussi hétérogènes que la France et la RDC ? Cette question interroge la robustesse de la découverte causale lorsque les données sont courtes et les structures économiques contrastées.

La deuxième question porte sur la quantification : le Double Machine Learning permet-il d'estimer un effet causal net de la formation brute de capital sur la croissance du PIB, statistiquement significatif et économiquement interprétable, après contrôle rigoureux des confondeurs macroéconomiques ? Cette question teste la capacité du CML à produire des estimations causales robustes sur données observationnelles.

La troisième question est comparative : l'effet causal de l'investissement sur la croissance diffère-t-il de manière structurelle et statistiquement significative entre pays développés et pays en développement ? Cette question vise à établir si deux régimes causaux distincts coexistent dans le panel, ou si un modèle universel peut rendre compte de la relation investissement-croissance.

4 Hypothèses de recherche

En réponse aux trois questions de recherche, trois hypothèses de travail sont formulées.

La première hypothèse est que les méthodes de Causal Discovery combinées en synthèse multi-méthodes produisent des structures causales cohérentes et robustes entre les 7 pays. Toutefois, le lien direct Formation de Capital vers PIB devrait être absent ou fragile dans la majorité des pays, car l'effet causal transite par des variables intermédiaires telles que le taux de change et la dette publique [Shojaie and Fox \(2022\)](#); [Runge et al. \(2019\)](#).

La deuxième hypothèse est que le Double Machine Learning, grâce au contrôle des confondeurs non linéaires via le RandomForest, révèle un effet causal net de l'investissement sur la croissance qui diffère significativement de la corrélation brute, et ce, en contradiction avec les résultats du test de Granger bivarié [Chernozhukov et al. \(2018\)](#); [Oprescu et al. \(2019\)](#).

La troisième hypothèse est que l'effet causal estimé est structurellement négatif dans les économies développées, en raison des rendements décroissants du capital et de l'effet d'éviction, et structurellement positif dans les économies africaines à déficit de capital, en raison de la rareté du capital physique. Cette hétérogénéité structurelle devrait être confirmée par un test statistique de différence entre groupes [Biaba et al. \(2025\)](#); [Mulomba et al. \(2025\)](#).

5 Objectifs

Ce stage vise à étudier l'impact causal de la Formation Brute de Capital sur la croissance économique (PIB), à partir de séries temporelles macroéconomiques annuelles. Des techniques avancées de Causal Machine Learning sont appliquées à un panel hétérogène de 7 pays (France, Allemagne, Maroc, Ghana, Nigeria, Angola, République Démocratique du Congo) sur la période 2000–2024.

Plus précisément, ce travail poursuit sept objectifs spécifiques. Le premier est d'harmoniser et préparer un dataset macroéconomique multi-sources (FMI WEO, FMI IFS, Banque Mondiale) couvrant 7 pays et 25 années (2000–2024), en appliquant les transformations nécessaires (log, winsorisation, standardisation) et la structure temporelle causale appropriée (variables en $t - 1$, outcome en t). Le deuxième est d'appliquer quatre méthodes de Causal Discovery (test de causalité de Granger, modèle VAR avec IRF et FEVD, modèle VECM avec test de Johansen, et algorithme PCMCi avec test CMiknn non linéaire) afin d'identifier les structures causales entre variables macroéconomiques dans chacun des 7 pays. Le troisième est de synthétiser les graphes causaux (DAGs) issus des 4 méthodes en un consensus multi-méthodes avec score de support, afin d'identifier les liens causaux robustes

versus fragiles. Le quatrième est d’estimer l’effet causal quantitatif de la Formation de Capital sur la croissance du PIB à l’aide du Double Machine Learning (LinearDML et CausalForestDML, econml), en contrôlant rigoureusement les 5 confondeurs macroéconomiques et les effets fixes pays. Le cinquième est d’analyser l’hétérogénéité des effets causaux entre pays et entre groupes à l’aide d’estimateurs d’effets de traitement hétérogènes (HTE). Le sixième est de comparer les performances des méthodes classiques d’économétrie (Granger, VAR, VECM) avec les méthodes de Causal Machine Learning (PCMCI, DML), afin d’évaluer les apports du CML. Le septième est de formuler des recommandations sur l’utilisation du CML en macroéconomie, notamment pour les contextes africains où les données sont limitées et les structures économiques hétérogènes.

6 Démarche méthodologique

La démarche méthodologique de ce travail s’articule en deux phases séquentielles et complémentaires, présentées en détail au Chapitre 2. La Phase 1, dite de *Causal Discovery*, vise à identifier les relations causales potentielles entre les sept variables macroéconomiques du panel. Elle combine quatre méthodes : le test de causalité de Granger, le modèle VAR (avec fonctions de réponse impulsionnelle et décomposition de la variance), le modèle VECM (avec test de cointégration de Johansen), et l’algorithme PCMCI avec test d’indépendance conditionnelle non linéaire CMiknn. Les résultats de ces quatre méthodes sont ensuite synthétisés en un graphe causal (DAG) de consensus multi-méthodes, assorti d’un score de support pour chaque lien identifié.

La Phase 2, fondée sur le *Double Machine Learning*, quantifie les effets causaux moyens (ATE) et hétérogènes (HTE) en contrôlant rigoureusement les confondeurs macroéconomiques. Deux estimateurs sont mobilisés : le LinearDML pour l’estimation de l’effet moyen global, par pays et par groupe de développement, et le CausalForestDML pour l’estimation des effets individuels hétérogènes. Les deux phases partagent les mêmes données prétraitées et produisent des résultats complémentaires, interprétés conjointement dans le Chapitre 3.

7 Contribution de l’étude

Ce travail apporte quatre contributions originales à la littérature. La contribution méthodologique consiste en un pipeline reproductible combinant Causal Discovery et Double Machine Learning avec une structure temporelle causale explicite (variables en $t - 1$, outcome en t). Ce pipeline est directement applicable à tout panel macroéconomique hétérogène. La contribution empirique est l’une des premières estimations comparatives, à notre connaissance, de l’effet causal de l’investissement sur la croissance couvrant simultanément deux pays européens développés et cinq pays africains sur 25 ans, avec des résultats statistiquement significatifs pour la France ($p = 0.023$) et la RDC ($p = 0.022$). La contribution économique est la mise en évidence d’une hétérogénéité structurelle hautement significative entre pays développés (ATE négatif) et pays en développement (ATE positif), confirmée par un test de Welch ($t = -6.97, p = 4.2 \times 10^{-9}$). La contribution pour l’Afrique est la démonstration de la faisabilité technique du CML sur données macroéconomiques africaines malgré les contraintes de séries courtes ($T = 25$), avec des adaptations méthodologiques spécifiques (CMiknn, cross-fitting adaptatif, effets fixes pays) [Biaba et al. \(2025\)](#); [Mulomba et al. \(2025\)](#).

8 Cadre institutionnel d'accueil

Le laboratoire ABIL (*Artificial intelligence, Big data and modeLing simulation*) est un centre de recherche interdisciplinaire fondé le 17 mars 2014 à Kinshasa, rattaché à la Faculté des Sciences de l'Université de Kinshasa (UNIKIN). Ses activités couvrent l'intelligence artificielle, le Big Data, la modélisation, la simulation et les systèmes décisionnels. Sa mission est de contribuer au développement de systèmes intelligents et décisionnels au service de la République Démocratique du Congo. Cette mission se concrétise par la recherche scientifique, la formation, la publication dans des revues spécialisées et la participation à des colloques nationaux et internationaux.

Depuis sa fondation, le Centre s'est constitué autour d'un noyau pluridisciplinaire réunissant des professeurs et chercheurs issus de plusieurs facultés de l'UNIKIN et d'universités partenaires (UNILU, UNIKIS). Son organisation suit un modèle fédératif à vocation nationale : les antennes locales disposent d'une autonomie opérationnelle tout en restant redevables envers le Centre central. Le siège est établi à la Faculté des Sciences de l'UNIKIN, bureau B5/E, Kinshasa.

ABIL est doté de trois organes statutaires : l'Assemblée Générale (organe suprême), le Conseil Scientifique (organe de contrôle) et le Conseil de Gestion (organe d'exécution). Le Tableau 1 présente la composition du bureau exécutif.

TABLE 1 – Bureau du Conseil de Gestion d'ABIL (Source ABIL)

Fonction	Titulaire	Grade
Directeur	Prof. Kasoro Mulenda Nathanael	Professeur, UNIKIN
Directeur Adjoint	Prof. Mabela Makengo Rostin	Professeur, UNIKIN
Secrétaire chargé des projets	Prof. Selain Kasereka Kabunga	Professeur, UNIKIN

Le présent travail a été réalisé dans ce cadre, sous la direction scientifique du Professeur Selain Kasereka Kabunga et la co-direction du Professeur Kasoro Mulenda Nathanaël, Directeur du laboratoire. L'encadrement académique et les ressources intellectuelles du Centre ont constitué un appui déterminant pour mener à bien cette recherche.

9 Structure du mémoire

Ce mémoire est organisé en quatre chapitres principaux, correspondant aux différentes étapes du travail réalisé.

Le Chapitre 1 intitulé « Cadre Théorique et État de l'Art » pose les fondements théoriques du Causal Machine Learning et dresse une revue de la littérature sur son application en macroéconomie. Les cadres de l'inférence causale (résultats potentiels, DAGs, do-calculus), les méthodes de Causal Discovery (Granger, VAR, VECM, PCMCI) et le Double Machine Learning y sont formalisés. L'état de l'art identifie les outils disponibles, les applications existantes et positionne notre travail par rapport à la littérature.

Le Chapitre 2 intitulé « Données et Méthodologie » décrit les données utilisées et la méthodologie mise en œuvre : sources, prétraitement, structure temporelle causale, pipeline d'analyse Phase 1 (Causal Discovery) et Phase 2 (Double Machine Learning).

Le Chapitre 3 intitulé « Résultats et Discussion » présente et analyse les résultats expérimentaux : structures causales identifiées (DAGs), effets causaux estimés (ATE global, par pays et par groupe),

hétérogénéité des effets (HTE) et comparaison des méthodes.

Le Chapitre 4 intitulé « Conclusion Générale et Perspectives » synthétise les apports scientifiques, les limites identifiées et les perspectives futures, notamment en vue de la rédaction d'un second article empirique.

CADRE THÉORIQUE ET ÉTAT DE L'ART

1.1 Introduction

Distinguer la causalité de la simple corrélation statistique est le défi central de l'économétrie causale. Ce chapitre est organisé en treize sections. La Section 1.2 présente les cadres théoriques de l'inférence causale. La Section 1.3 formalise la causalité de Granger et ses limites. La Section 1.4 expose le modèle VAR et l'analyse des dynamiques. La Section 1.5 détaille la cointégration et le modèle VECM. La Section 1.6 présente l'algorithme PCMCI pour la découverte causale non linéaire. La Section 1.7 formalise le Double Machine Learning. La Section 1.8 analyse les limites des modèles économétriques classiques. La Section 1.9 retrace l'émergence du Causal Machine Learning en macroéconomie. La Section 1.10 recense les outils, bibliothèques et algorithmes du CML. La Section 1.11 examine les applications en contexte africain. La Section 1.12 examine les travaux sur l'effet de l'investissement sur la croissance. La Section 1.13 positionne notre travail dans l'état de l'art. La Section 1.14 conclut le chapitre.

1.2 Inférence Causale

L'inférence causale repose sur deux paradigmes complémentaires qui structurent l'ensemble des méthodes employées dans ce travail.

1.2.1 Le Framework des Résultats Potentiels

Introduit par Rubin (2007), le framework des résultats potentiels formule l'effet causal moyen (ATE) selon l'Équation (1.1).

$$ATE = \mathbb{E}[Y(1) - Y(0)] \quad (1.1)$$

$Y(1)$ désigne le résultat sous traitement et $Y(0)$ le résultat en l'absence de traitement. Dans notre contexte, le traitement T correspond à la Formation de Capital en $t - 1$ et l'outcome Y à la croissance du PIB en t . Ce cadre est adapté à l'estimation des effets de politique économique lorsque la randomisation est irréalisable Guo et al. (2020); Oprescu et al. (2019).

1.2.2 Les Modèles Causaux Graphiques (DAGs)

Le second paradigme, fondé sur Pearl (2009), utilise les graphes acycliques dirigés (DAGs) pour représenter les relations causales entre variables. L'opérateur *do-calculus* distingue l'observation de l'intervention : l'Équation (1.2) exprime la distribution de Y lorsqu'on intervient activement sur X .

$$P(Y \mid do(X = x)) \quad (1.2)$$

Cette expression capture la distribution de Y sous intervention active, à distinguer de la simple observation conditionnelle $P(Y | X = x)$ [Guo et al. \(2020\)](#). Dans notre travail, les DAGs sont construits par synthèse multi-méthodes (Granger, VAR, PCMCI) et permettent d’identifier les chemins causaux entre les sept variables macroéconomiques du panel.

1.3 Causalité de Granger et Extensions

1.3.1 Test de Causalité de Granger

La causalité de Granger est le point de départ historique de l’analyse causale sur séries temporelles en économie. Une variable X cause au sens de Granger une variable Y si ses valeurs passées améliorent la prédiction de Y au-delà de ce qu’apportent seules les valeurs passées de Y [Shojaie and Fox \(2022\)](#). L’Équation (1.3) formalise ce test :

$$Y_t = \sum_{i=1}^p \alpha_i Y_{t-i} + \sum_{j=1}^q \beta_j X_{t-j} + \epsilon_t \quad (1.3)$$

p et q sont les ordres de lag maximaux, α_i et β_j les coefficients de régression, et ϵ_t le terme d’erreur bruit blanc. X cause Y au sens de Granger si au moins un $\beta_j \neq 0$, testé par un test F [Shojaie and Fox \(2022\)](#); [Biaba et al. \(2025\)](#).

Le test est appliqué à toutes les paires de variables pour chacun des 7 pays, avec un lag maximal de 3 et un seuil $\alpha = 0.05$. La p-value minimale sur les lags 1 à 3 est retenue pour chaque paire.

1.3.2 Limites du Test de Granger

Le test de Granger souffre de deux limites structurelles pour notre contexte. Sa formulation classique est fondamentalement **linéaire** et **bivariée** : elle ne contrôle pas l’influence simultanée des autres variables confondantes [Shojaie and Fox \(2022\)](#). Un test peut ainsi détecter une relation entre le taux d’intérêt et l’inflation sans isoler l’effet causal de l’un sur l’autre en présence d’un troisième facteur commun. Ces lacunes justifient le recours aux méthodes VAR, VECM et PCMCI.

1.4 Modèle VAR et Analyse des Dynamiques

1.4.1 Définition et Formulation

Le modèle VAR (*Vector Autoregression*) généralise le test de Granger à un système multivarié où chaque variable est modélisée comme une fonction linéaire de ses propres valeurs passées et des valeurs passées de toutes les autres variables du système [Maitra \(2025\)](#). Pour un vecteur de k variables $\mathbf{Y}_t$, le VAR(p) s’écrit selon l’Équation (1.4) :

$$\mathbf{Y}_t = \mathbf{A}_1 \mathbf{Y}_{t-1} + \mathbf{A}_2 \mathbf{Y}_{t-2} + \dots + \mathbf{A}_p \mathbf{Y}_{t-p} + \epsilon_t \quad (1.4)$$

$\mathbf{A}_i$ sont des matrices de coefficients de dimension $k \times k$ et ϵ_t un vecteur de termes d’erreur. Le nombre de lags p est sélectionné par minimisation des critères d’information AIC, BIC, HQIC et FPE [Maitra \(2025\)](#).

Le VAR est estimé séparément pour chacun des 7 pays sur les données transformées et laggées, avec un lag maximal initial de 4. Le critère BIC est retenu pour la sélection, conduisant à un lag optimal de 2 pour 6 pays sur 7 (lag 1 pour la RDC).

1.4.2 Fonction de Réponse Impulsionnelle (IRF)

La fonction de réponse impulsionnelle (IRF) mesure la réponse dynamique d’une variable Y à un choc unitaire sur une variable X au fil du temps, toutes les autres variables étant maintenues constantes. Elle permet de visualiser la transmission des chocs macroéconomiques dans le système [Maitra \(2025\)](#). Notre analyse porte particulièrement sur la réponse de GDP_Growth , le taux de croissance annuel du PIB réel, exprimé en pourcentage à un choc sur $Capital_Formation_lag1$, l’investissement total rapporté au PIB et décalé d’une année. Ces deux variables, ainsi que l’ensemble du jeu de données, sont décrites au Tableau [2.1](#).

1.4.3 Décomposition de la Variance (FEVD)

La décomposition de la variance de l’erreur de prévision (FEVD) mesure la proportion de la variance d’une variable expliquée par les chocs de chacune des autres variables du système, à différents horizons temporels. Elle quantifie la contribution relative de chaque variable à la dynamique de GDP_Growth , le taux de croissance annuel du PIB réel exprimé en pourcentage, qui constitue la variable expliquée de notre étude (cf. Tableau [2.1](#)) [Maitra \(2025\)](#).

1.5 Cointégration et Modèle VECM

1.5.1 Test de Cointégration de Johansen

Des séries temporelles non stationnaires en niveau mais stationnaires en différence (intégrées d’ordre 1, notées $I(1)$) peuvent partager une relation d’équilibre de long terme appelée cointégration [Chen et al. \(2021\)](#). Le test de Johansen détermine le nombre de relations de cointégration (rang r) au sein d’un système multivarié, en comparant la statistique de trace λ_{trace} aux valeurs critiques. L’Équation (1.5) en donne la formulation :

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i) \quad (1.5)$$

T est le nombre d’observations, k le nombre de variables et $\hat{\lambda}_i$ les valeurs propres estimées [Chen et al. \(2021\)](#). Le test est appliqué à chacun des 7 pays au seuil de significativité de 95%.

1.5.2 Modèle VECM

Le modèle VECM (*Vector Error Correction Model*) est une reformulation du VAR adaptée aux variables cointégrées. Sa dynamique se décompose en deux composantes : les ajustements de court terme et la correction vers l’équilibre de long terme [Chen et al. \(2021\)](#); [Maitra \(2025\)](#). L’Équation (1.6) en donne la formulation générale :

$$\Delta Y_t = \alpha \beta' Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \epsilon_t \quad (1.6)$$

β est la matrice des vecteurs de cointégration (relations de long terme) et α la matrice des vitesses d'ajustement vers l'équilibre. Un coefficient α_i proche de zéro pour une variable indique qu'elle ne corrige pas les déséquilibres : elle est exogène dans le système.

Ce résultat est central : le coefficient d'ajustement de `Capital_Formation` vaut $\alpha \approx 0.0003$, ce qui confirme son rôle de variable exogène et renforce la validité de son utilisation comme variable de traitement dans le DML.

1.6 Découverte Causale Non Linéaire : PCMCI

1.6.1 Principe et Motivation

Les méthodes Granger, VAR et VECM reposent sur des hypothèses de linéarité qui peuvent être inadéquates aux dynamiques macroéconomiques complexes, notamment dans les économies africaines émergentes. L'algorithme PCMCI (*Peter and Clark Momentary Conditional Independence*) [Runge et al. \(2019\)](#) surmonte ces limites en combinant deux étapes : une phase de sélection des parents causaux (PC) et un test d'indépendance conditionnelle momentanée (MCI). Cette architecture permet de détecter des relations causales non linéaires tout en contrôlant les faux positifs dans les systèmes à haute dimension [Runge et al. \(2019\)](#); [Krich et al. \(2020\)](#).

1.6.2 Formulation

Pour une série temporelle multivariée $\mathbf{X} = (X^1, \dots, X^k)$, PCMCI teste si la variable $X_{t-\tau}^i$ cause X_t^j conditionnellement à l'ensemble des parents causaux de X_t^j . L'Équation (1.7) formalise ce test d'indépendance conditionnelle :

$$X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \mathcal{P}(X_t^j) \setminus \{X_{t-\tau}^i\} \quad (1.7)$$

$\mathcal{P}(X_t^j)$ désigne l'ensemble des parents causaux de X_t^j et τ le lag temporel considéré [Runge et al. \(2019\)](#). Un lien causal est retenu si l'hypothèse nulle d'indépendance conditionnelle est rejetée au seuil $\alpha = 0.05$, après correction FDR de Benjamini-Hochberg pour les tests multiples.

1.6.3 Test CMiknn

Le test d'indépendance retenu est celui fondé sur l'information mutuelle conditionnelle avec k plus proches voisins (CMiknn, $k = 5$) [Runge et al. \(2019\)](#). Ce test est robuste aux non-linéarités et ne suppose aucune forme paramétrique. Il est particulièrement adapté aux séries macroéconomiques africaines qui présentent des distributions asymétriques et des queues épaisses, comme en témoigne le skewness de `Capital_Formation` ($= -1.20$) mis en évidence lors de l'analyse descriptive.

Les paramètres retenus sont : $\tau_{min} = 1$, $\tau_{max} = 3$, $\alpha_{PC} = 0.05$, correction FDR de Benjamini-Hochberg. PCMCI opère sur les données **non lagguées**, l'algorithme gérant lui-même la structure temporelle des décalages [Runge et al. \(2019\)](#).

1.7 Double Machine Learning

1.7.1 Motivation et Principe

Le Double Machine Learning (DML), introduit par [Chernozhukov et al. \(2018\)](#), est une méthode d'estimation causale semi-paramétrique. Elle combine la flexibilité des algorithmes d'apprentissage automatique avec la rigueur de l'inférence causale. Son objectif est d'estimer l'effet causal moyen (ATE) d'un traitement continu T sur un outcome Y , en présence de confondeurs potentiellement nombreux et non linéaires X , à partir de données observationnelles [Oprescu et al. \(2019\)](#).

Deux observations empiriques motivent le recours au DML dans notre contexte. La corrélation brute entre `Capital_Formation`, l'investissement total rapporté au PIB, qui constitue notre variable de traitement et `GDP_Growth`, le taux de croissance annuel du PIB réel, notre variable expliquée (cf. Tableau 2.1) est quasi nulle ($r = -0.02$), ce qui indique que l'effet causal est masqué par les confondeurs. Par ailleurs, le coefficient de détermination $R^2 = 0.772$ pour la prédiction du traitement à partir des confondeurs confirme que l'investissement est fortement confondu : sans contrôle rigoureux, tout estimateur naïf serait sévèrement biaisé.

1.7.2 Formulation du Modèle Partiellement Linéaire

Le DML repose sur le modèle partiellement linéaire de [Robinson \(1988\)](#). Les Équations (1.8) et (1.9) présentent respectivement le modèle de résultat et le modèle de traitement :

$$Y = \theta T + g(X) + \epsilon \quad (1.8)$$

$$T = m(X) + \eta \quad (1.9)$$

θ est l'ATE à estimer, $g(X)$ et $m(X)$ sont des fonctions arbitraires (potentiellement non linéaires) des confondeurs X , et ϵ, η sont des termes d'erreur [Chernozhukov et al. \(2018\)](#). L'idée centrale est de résidualiser Y et T par rapport à X via des modèles ML, puis d'estimer θ par régression des résidus. L'Équation (1.10) formalise cette étape :

$$\tilde{Y} = Y - \hat{g}(X), \quad \tilde{T} = T - \hat{m}(X) \quad (1.10)$$

L'Équation (1.11) donne l'estimateur de l'effet causal moyen :

$$\hat{\theta} = \frac{\text{Cov}(\tilde{Y}, \tilde{T})}{\text{Var}(\tilde{T})} \quad (1.11)$$

1.7.3 Cross-Fitting

Pour éviter le biais de surapprentissage (*overfitting bias*), le DML utilise la technique de *cross-fitting* : les données sont divisées en K plis ; les modèles nuisances $\hat{g}$ et $\hat{m}$ sont entraînés sur $K - 1$ plis et appliqués sur le pli restant [Chernozhukov et al. \(2018\)](#). Nous retenons $K = 5$ plis avec RandomForest (400 arbres, profondeur maximale 6). Ce choix est justifié par sa performance supérieure ($R_Y^2 = 0.392$, $R_T^2 = 0.772$) face aux alternatives linéaires (Ridge : $R_Y^2 = 0.221$; Lasso : $R_Y^2 = -0.045$) [Oprescu et al. \(2019\)](#).

1.7.4 Causal Forest et Effets Hétérogènes

L'estimation des effets de traitement hétérogènes (HTE) requiert d'aller au-delà de l'ATE global. Nous utilisons le *CausalForestDML* [Oprescu et al. \(2019\)](#), une extension du Double Machine Learning qui estime un effet individuel $\tau(x)$ pour chaque observation. L'Équation (1.12) définit cet effet :

$$\tau(x) = \mathbb{E}[Y(1) - Y(0) \mid X = x] \quad (1.12)$$

Le Causal Forest partitionne l'espace des confondeurs en sous-groupes homogènes et estime un ATE local dans chacun d'eux. Cette granularité permet d'identifier quels pays ou contextes macroéconomiques bénéficient davantage d'une hausse de l'investissement [Oprescu et al. \(2019\)](#); [Zhao and Liu \(2023\)](#).

1.7.5 Hypothèses d'Identification

La validité des estimations DML repose sur trois hypothèses d'identification [Chernozhukov et al. \(2018\)](#). L'hypothèse d'**Unconfoundedness** stipule que, conditionnellement aux confondeurs X , le traitement T est indépendant de l'outcome potentiel $Y(t)$; le décalage temporel d'un an renforce cette hypothèse en réduisant le risque de simultanéité. L'hypothèse d'**Overlap** requiert que chaque unité ait une probabilité non nulle de recevoir n'importe quelle valeur du traitement, soit $0 < P(T \mid X) < 1$. L'hypothèse **SUTVA** (*Stable Unit Treatment Value Assumption*) suppose que l'outcome d'un pays ne dépend pas du niveau d'investissement des autres pays.

1.7.6 Spécification du Modèle dans Notre Travail

Le modèle DML est spécifié comme suit. L'outcome Y est $\text{GDP_Growth}(t)$, le taux de croissance annuel du PIB réel en année t . Le traitement T est $\text{Capital_Formation}(t-1)$, la formation brute de capital (% du PIB) en année $t - 1$. Les confondeurs X comprennent cinq variables macroéconomiques en $t - 1$ (*Inflation*, *Government_Debt*, *Trade_Balance*, *Exchange_Rate*, *Reserves*), auxquelles s'ajoutent des effets fixes pays (*country dummies*) pour contrôler l'hétérogénéité non observée spécifique à chaque pays.

Le décalage d'une année entre le traitement et l'outcome ($t - 1 \rightarrow t$) répond à deux exigences complémentaires. Il impose d'abord une direction temporelle claire : l'investissement de l'année $t - 1$ ne peut logiquement pas être causé par la croissance de l'année t . Il est ensuite cohérent avec la théorie keynésienne de l'accélérateur, selon laquelle les effets de l'investissement sur la production se manifestent avec un délai [Chernozhukov et al. \(2018\)](#).

Ce bilan théorique met en évidence trois enseignements structurants. Les méthodes classiques (Granger, VAR, VECM) identifient les structures dynamiques entre variables macroéconomiques, mais leur hypothèse de linéarité et leur incapacité à contrôler l'ensemble des confondeurs simultanément en limitent la portée causale [Shojaie and Fox \(2022\)](#); [Maitra \(2025\)](#). L'algorithme PCMCi [Runge et al. \(2019\)](#) lève la contrainte de linéarité grâce au test CMiknn non paramétrique, mais sa puissance statistique reste faible sur des séries courtes ($T = 25$ observations par pays), ce qui rend indispensable sa combinaison avec les méthodes classiques dans une synthèse multi-méthodes. Le Double Machine Learning [Chernozhukov et al. \(2018\)](#); [Oprescu et al. \(2019\)](#) est la seule méthode produisant un effet causal quantifié assorti d'un intervalle de confiance, tout en contrôlant rigoureusement les confondeurs

non linéaires ; c’est pourquoi il constitue le cœur de notre contribution empirique.

1.8 Limites des Modèles Économétriques Classiques

Les modèles économétriques traditionnels ont longtemps occupé une place centrale dans l’analyse macroéconomique. Les modèles VAR et VECM capturent les dynamiques de court et long terme entre variables macroéconomiques [Maitra \(2025\)](#); [Chen et al. \(2021\)](#). Les modèles DSGE (*Dynamic Stochastic General Equilibrium*) fournissent un cadre structurel pour l’évaluation des politiques monétaires et fiscales.

Ces approches souffrent cependant de limites structurelles bien documentées [Shojaie and Fox \(2022\)](#); [Guo et al. \(2020\)](#); [Biaba et al. \(2025\)](#). L’hypothèse de linéarité exclut les effets de seuil, les non-linéarités et les interactions complexes fréquemment observées dans les économies émergentes. L’hypothèse de stationnarité impose des transformations préalables (différenciation, log) qui peuvent masquer des relations causales de long terme. Ces modèles mesurent des associations statistiques et non des relations causales au sens de l’intervention : un VAR peut identifier que l’inflation précède la croissance sans établir que l’une cause l’autre [Shojaie and Fox \(2022\)](#). En présence de nombreuses variables confondantes et d’interactions non linéaires, les modèles classiques ne permettent pas de contrôler simultanément tous les biais de confusion [Guo et al. \(2020\)](#). Enfin, dans les économies subsahariennes où les séries temporelles sont courtes et les données manquantes fréquentes, les hypothèses des modèles classiques sont encore moins tenables [Mulomba et al. \(2025\)](#); [Biaba et al. \(2025\)](#). Ces limites ont ouvert la voie au Causal Machine Learning, qui s’impose comme une alternative complémentaire aux approches classiques [Guo et al. \(2020\)](#).

1.9 Émergence du Causal Machine Learning en Macroéconomie

1.9.1 Du Prédicatif au Causal

L’essor du Machine Learning en économie a d’abord été dominé par des approches purement prédictives, cherchant à optimiser la précision des prévisions macroéconomiques sans se préoccuper de l’interprétation causale des modèles [Guo et al. \(2020\)](#). À partir de 2018, la formalisation du Double Machine Learning par [Chernozhukov et al.](#) a marqué un tournant décisif. Le développement de bibliothèques dédiées comme EconML [Oprescu et al. \(2019\)](#) et CausalML [Zhao and Liu \(2023\)](#) a consolidé le CML comme champ de recherche autonome, combinant explicitement la puissance prédictive du ML avec la rigueur de l’inférence causale.

1.9.2 Applications en Macroéconomie

Les applications du CML en macroéconomie demeurent peu nombreuses, bien qu’elles couvrent plusieurs domaines [Biaba et al. \(2025\)](#). Des approches récentes extraient la structure causale sous-jacente aux modèles DSGE à partir de données réelles, contournant les hypothèses paramétriques fortes des modèles traditionnels [Runge et al. \(2019\)](#). La découverte causale a également été employée pour établir des dépendances temporelles entre variables macroéconomiques, environnementales et socio-politiques, ouvrant la voie à une modélisation multi-système [Runge et al. \(2019\)](#). Des études récentes montrent que le CML peut compléter les approches économétriques classiques, notamment

pour l'estimation d'effets hétérogènes et la robustesse en contexte observationnel [Oprescu et al. \(2019\)](#); [Guo et al. \(2020\)](#). Par ailleurs, [Cordoni and Sancetta \(2024\)](#) développent un cadre causal pour séries temporelles à haute dimension et illustrent des cas d'usage macroéconomiques, dont l'impact des chocs pétroliers sur l'économie.

1.9.3 Adoption Fragmentée et Gap Africain

Malgré ce potentiel, l'adoption du CML en macroéconomie reste fragmentée et très inégalement répartie dans le monde [Biaba et al. \(2025\)](#). Moins de 1% des études combinant CML et macroéconomie intègrent une analyse spatiotemporelle, dimension pourtant essentielle pour l'évaluation des politiques publiques dans des contextes géographiquement diversifiés. Les applications dans les pays africains sont quasi inexistantes : les rares travaux disponibles, comme celui de [Mulomba et al. \(2025\)](#), soulignent que l'application du CML reste entravée par le manque de séries longues, la rareté des contrefactuels plausibles et la difficulté de modéliser l'économie informelle. Enfin, les outils existants (DoWhy, EconML, CausalML) sont conçus pour des environnements institutionnels, linguistiques et technologiques qui ne correspondent pas aux réalités des pays africains [Biaba et al. \(2025\)](#).

1.10 Outils, Bibliothèques et Algorithmes du CML

L'essor du Causal Machine Learning au cours de la dernière décennie a conduit au développement d'un écosystème riche de bibliothèques logicielles et d'algorithmes permettant d'estimer des effets causaux à partir de données observationnelles. Le Tableau 1.1 résume les principaux outils disponibles, et le Tableau 1.2 compare les principaux algorithmes de découverte causale pour séries temporelles.

DoWhy propose une analyse causale en quatre étapes (modélisation, identification, estimation, réfutation) avec modélisation graphique et tests de robustesse. EconML, développée par Microsoft Research, implémente les méta-learners (S/T/X) et le Double Machine Learning pour l'économétrie [Oprescu et al. \(2019\)](#). CausalML, issue de Uber Research, se spécialise dans le *uplift modeling* et l'estimation des effets de traitement hétérogènes (CATE) [Zhao and Liu \(2023\)](#).

En matière de découverte causale, trois algorithmes dominent. PCMCI (*Peter and Clark Momentary Conditional Independence*) est conçu pour les séries temporelles à haute dimension [Runge et al. \(2019\)](#). FCI (*Fast Causal Inference*) gère les systèmes avec variables latentes [Strobl et al. \(2018\)](#). Les réseaux bayésiens permettent l'intégration de connaissances a priori dans la modélisation causale [Liu et al. \(2019\)](#). Malgré leur puissance, ces outils rencontrent des obstacles significatifs dans les pays africains : absence de longues séries temporelles, infrastructure informatique limitée, et manque d'expertise dans ces technologies émergentes. Il est donc crucial de promouvoir des outils open-source accessibles et bien documentés [Mulomba et al. \(2025\)](#).

TABLE 1.1 – Principaux outils et méthodes du CML pour les séries temporelles macroéconomiques

Outil / Méthode	Caractéristiques principales	Référence
DoWhy	Graphes causaux, contrefactuels, support séries temporelles	Guo et al. (2020)
EconML	Meta-learners (S/T/X), Double ML pour l'économétrie	Oprescu et al. (2019)
CausalML	Uplift modeling, forêts causales, réseaux de neurones	Zhao and Liu (2023)
PCMCI, FCI, Réseaux bayésiens	Découverte causale en séries temporelles	Runge et al. (2019)
Deep Learning Causal	GNNs, autoencodeurs pour non-linéarités	Guo et al. (2020)

TABLE 1.2 – Comparaison des principaux algorithmes de découverte causale pour séries temporelles

Algorithme	Principe	Forces	Limites
PCMCI	Indépendance conditionnelle momentanée	Gère les séries haute dimension, réduit les faux positifs	Nécessite de grands échantillons
FCI	Ind. conditionnelle avec variables latentes	Gère les variables cachées, graphes partiellement orientés	Coût computationnel élevé
Réseaux bayésiens	Modèles graphiques probabilistes	Flexibilité, encode les connaissances a priori	Sensible aux hypothèses a priori

1.11 Applications en Contexte Africain

L'application du CML en macroéconomie reste encore marginale et inégalement répartie. Les travaux récents se concentrent sur trois domaines : l'analyse structurelle des modèles DSGE, la détection de causalités entre agrégats économiques, et l'évaluation de politiques publiques à partir de données observationnelles.

Dans le contexte africain, les obstacles sont particulièrement prononcés. [Mulomba et al. \(2025\)](#) identifient trois catégories de contraintes : la parcimonie temporelle des données, l'hétérogénéité des sources, et la prédominance de systèmes informels non captés par les statistiques officielles. Malgré cela, des initiatives émergentes en Ghana, au Sénégal et au Rwanda montrent que des approches localement ancrées permettent une meilleure appropriation des outils [Olayungbo \(2019\)](#).

Face à ces défis, cinq piliers d'un cadre d'adaptation ont été proposés par [Biaba et al. \(2025\)](#) : le diagnostic contextuel, la sélection adaptative des outils, l'adaptation méthodologique (méthodes bayésiennes, transfer learning), la validation collaborative avec les experts locaux, et l'implémentation progressive pour construire les compétences locales. Ces principes guident les choix méthodologiques de notre propre travail empirique.

1.12 Travaux sur l'Effet de l'Investissement sur la Croissance

1.12.1 Perspective Théorique

La relation entre investissement et croissance économique est l'une des plus étudiées en économie du développement. Depuis le modèle de [Solow \(1956\)](#), la théorie prédit un effet positif de l'accumulation du capital physique sur la production, sous l'hypothèse de rendements décroissants. Les modèles de croissance endogène ([Romer, 1990](#); [Lucas, 1988](#)) ont enrichi cette perspective en intégrant le capital humain et le progrès technologique comme moteurs complémentaires. Cette relation théorique positive est loin d'être universelle dans les données empiriques. Plusieurs mécanismes peuvent inverser ou neutraliser l'effet de l'investissement sur la croissance [Mazumder et al. \(2025\)](#). Dans les économies à stock de capital élevé, chaque unité supplémentaire d'investissement génère un rendement marginal décroissant, et l'effet net sur la croissance peut ainsi devenir faible, voire négatif. L'investissement public financé par la dette peut évincer l'investissement privé en augmentant les taux d'intérêt réels, réduisant l'effet net sur la croissance (*crowding-out*) [Mazumder et al. \(2025\)](#). Dans les économies dépendantes des matières premières comme le Nigeria ou l'Angola, les flux d'investissement sont souvent concentrés dans le secteur extractif et génèrent peu d'effets d'entraînement sur le reste de l'économie (*resource curse*). Enfin, dans les économies à gouvernance faible comme la RDC, l'investissement supplémentaire peut être détourné ou mal alloué, réduisant son impact sur la croissance réelle [Mulomba et al. \(2025\)](#).

1.12.2 Études Empiriques sur Données de Panel Africain

Les études empiriques sur l'effet de l'investissement en Afrique subsaharienne donnent des résultats contrastés. [Olayungbo \(2019\)](#) applique les réseaux bayésiens à l'analyse de la politique monétaire au Nigeria et montre que les relations entre variables macroéconomiques sont fortement non linéaires et dépendent du régime conjoncturel.

[Mulomba et al. \(2025\)](#) identifient trois obstacles récurrents à l'application du CML sur données subsahariennes : la parcimonie des séries disponibles, l'hétérogénéité structurelle entre pays, et la prévalence de l'économie informelle, absente des indicateurs officiels.

La littérature économétrique de panel sur la relation investissement-croissance en Afrique subsaharienne offre des résultats nuancés. [Bond et al. \(2001\)](#) montrent que les estimateurs GMM dynamiques, appliqués à des panels africains, produisent des coefficients d'investissement significatifs mais fortement sensibles à la spécification des instruments [Bond et al. \(2001\)](#). [Calderón and Servén \(2008\)](#) documentent que l'impact de l'investissement sur la croissance en Afrique subsaharienne est conditionnel à la qualité des infrastructures : un dollar investi dans les économies à faible dotation infrastructurelle génère un rendement marginal plus élevé que dans les économies matures [Calderón and Servén \(2008\)](#). Ces travaux convergent vers un constat que notre étude cherche à tester empiriquement à l'aide du CML : l'effet de l'investissement sur la croissance n'est pas linéaire et dépend du contexte macroéconomique et institutionnel de chaque pays. Notre approche par Double Machine Learning, en estimant des effets hétérogènes par pays plutôt qu'un coefficient moyen de panel, permet de capturer cette hétérogénéité structurelle que les estimateurs GMM classiques ne peuvent identifier qu'imparfaitement.

Ces résultats convergent vers un constat central : **il n'existe pas d'effet universel de l'investis-**

sement sur la croissance. L'effet dépend du niveau de développement, du stock de capital existant, de la qualité institutionnelle et du contexte macroéconomique global. Ce constat justifie l'approche hétérogène adoptée dans notre travail, qui estime des effets distincts par pays et par groupe de développement.

1.13 Positionnement de Notre Travail dans l'État de l'Art

1.13.1 Ce que la Littérature a Établi

Trois acquis scientifiques se dégagent de la revue conduite dans ce chapitre. Les méthodes de Causal Discovery (Granger, VAR, VECM, PCMCI) sont complémentaires, et leur combinaison en un consensus multi-méthodes produit des structures causales plus robustes que chaque méthode prise isolément [Runge et al. \(2019\)](#); [Shojaie and Fox \(2022\)](#). Le Double Machine Learning est la méthode la plus adaptée à l'estimation d'effets causaux quantifiés en présence de confondeurs non linéaires et de données observationnelles [Chernozhukov et al. \(2018\)](#); [Oprescu et al. \(2019\)](#). L'application du CML en Afrique subsaharienne est encore embryonnaire et se heurte à des contraintes spécifiques de données et d'infrastructure [Mulomba et al. \(2025\)](#); [Biaba et al. \(2025\)](#).

1.13.2 Le Gap que Notre Travail Comble

Aucune étude publiée, à notre connaissance, ne réunit simultanément les caractéristiques suivantes : l'application conjointe du Causal Discovery (4 méthodes) et du Double Machine Learning sur un même panel macroéconomique ; un panel multi-pays hétérogène combinant économies développées (France, Allemagne) et économies subsahariennes (Angola, Ghana, Nigeria, RDC) avec un pays intermédiaire (Maroc) ; une période longue 2000–2024 (25 ans) avec données harmonisées multi-sources (FMI WEO, FMI IFS, Banque Mondiale) ; une synthèse multi-méthodes des graphes causaux avec score de support quantifié ($s \in \{1, 2, 3\}$) ; et l'estimation des effets hétérogènes (HTE) par pays et par groupe de développement avec test statistique de différence.

1.13.3 Notre Contribution Spécifique

Notre travail apporte quatre contributions originales à la littérature méthodologique, empirique, économique et régionale détaillées en Section 7 de l'Introduction Générale et synthétisées dans la Conclusion.

1.14 Conclusion partielle

Ce chapitre a présenté les fondements théoriques et l'état de l'art qui guident notre démarche analytique. Trois conclusions en ressortent. Le CML est encore marginal en macroéconomie : malgré son potentiel démontré pour l'estimation causale en présence de confondeurs non linéaires, son adoption reste fragmentée, et en Afrique subsaharienne, moins de 1% des études l'appliquent à des données macroéconomiques [Biaba et al. \(2025\)](#). L'effet de l'investissement sur la croissance n'est pas universel : la littérature empirique montre que cet effet dépend du niveau de développement, du stock de capital existant, de la qualité institutionnelle et de la structure économique du pays [Mazumder](#)

[et al. \(2025\)](#); [Mulomba et al. \(2025\)](#); [Olayungbo \(2019\)](#), ce qui justifie une approche pays par pays. Notre travail comble un gap identifié : en combinant Causal Discovery et Double Machine Learning sur un panel hétérogène de 7 pays sur 25 ans (2000–2024), nous répondons directement aux lacunes identifiées dans [Biaba et al. \(2025\)](#).

Le chapitre suivant décrit les données utilisées et la méthodologie mise en œuvre pour produire cette contribution empirique.

DONNÉES ET MÉTHODOLOGIE

2.1 Introduction

Ce chapitre décrit l'ensemble du pipeline mis en œuvre pour produire les résultats empiriques de ce travail. La Section 2.2 présente les sources de données utilisées. La Section 2.3 décrit les étapes de prétraitement. La Section 2.4 expose la structure temporelle causale retenue. La Section 2.5 détaille la Phase 1 de Causal Discovery. La Section 2.6 présente la Phase 2 de Double Machine Learning. La Section 2.7 synthétise les choix méthodologiques clés.

2.2 Sources de Données

2.2.1 Panel de Pays et Période

Le dataset couvre 7 pays répartis en deux groupes de développement sur 25 ans (2000–2024), pour un total de 175 observations avant lagging (7×25). Le groupe des pays développés comprend la France et l'Allemagne. Le groupe des pays en développement comprend l'Angola, le Ghana, le Nigeria, le Maroc et la République Démocratique du Congo (RDC).

2.2.2 Sources Primaires

Les données proviennent de trois sources officielles harmonisées. Le FMI WEO (*World Economic Outlook*) fournit cinq variables extraites via trois fichiers WEO régionaux (Europe, Afrique subsaharienne, Maroc) : GDP_Growth (code NGDP_RPCH), Capital_Formation (code NID_NGDP), Inflation (code PCPIPCH), Government_Debt (code GGXWDG_NGDP), Trade_Balance (code BCA_NGDPD). Le FMI IFS (*International Financial Statistics*) fournit deux variables extraites via des fichiers IFS séparés pour l'Afrique et la zone Euro : Exchange_Rate (indicateur : *Domestic currency per US Dollar*) et Reserves (indicateur : *Reserves excluding gold, foreign exchange*). La Banque Mondiale est utilisée pour la vérification et la complétion des données WEO lorsque nécessaire.

2.2.3 Les 7 Variables Macroéconomiques

Le Tableau 2.1 présente les 7 variables retenues avec leur unité, leur source et leur rôle dans l'analyse¹ :

1. L'indicateur NID_NGDP du FMI mesure l'investissement total, défini comme la somme de la formation brute de capital fixe, des variations de stocks et des acquisitions nettes d'objets de valeur. Il ne se limite donc pas à la formation brute de capital fixe au sens strict.

TABLE 2.1 – Description des variables macroéconomiques utilisées

Variable	Unité	Source	Rôle
GDP_Growth	% annuel	FMI WEO	Outcome Y
Capital_Formation	% du PIB	FMI WEO	Traitement T
Inflation	% annuel	FMI WEO	Confondeur X
Government_Debt	% du PIB	FMI WEO	Confondeur X
Trade_Balance	% du PIB	FMI WEO	Confondeur X
Exchange_Rate	USD	FMI IFS	Confondeur X
Reserves	USD	FMI IFS	Confondeur X

Note sur le taux de change : la zone euro partageant une monnaie commune, le taux de change EUR/USD a été dupliqué identiquement pour la France et l’Allemagne à partir du fichier IFS zone Euro.

2.2.4 Architecture du système

La Figure 2.1 schématise l’architecture générale du pipeline d’analyse causale mis en œuvre dans ce travail. Ce pipeline s’organise en deux phases séquentielles. La première phase, dite de *Causal Discovery*, vise à identifier les relations causales potentielles entre variables macroéconomiques à partir de quatre méthodes complémentaires : le test de causalité de Granger, le modèle VAR, le modèle VECM et l’algorithme PCMCI. La seconde phase, fondée sur le *Double Machine Learning*, quantifie les effets causaux moyens (ATE) et hétérogènes (HTE) en contrôlant rigoureusement les confondeurs macroéconomiques. Les deux phases partagent les mêmes données prétraitées et produisent des résultats complémentaires qui sont ensuite interprétés conjointement.

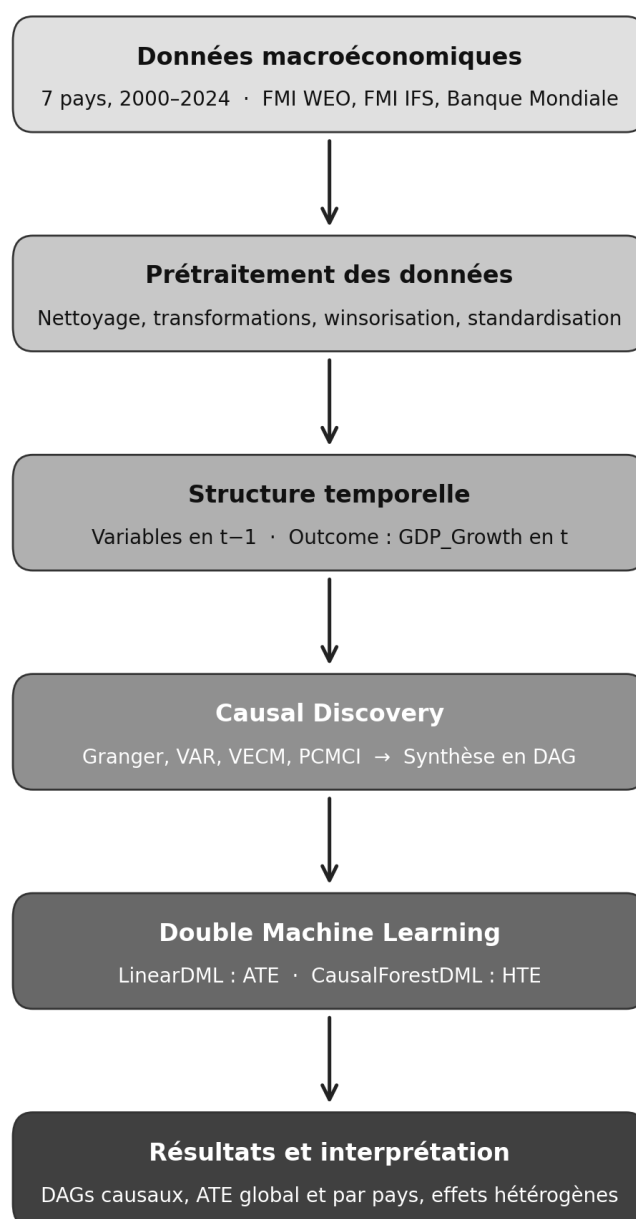


FIGURE 2.1 – Architecture séquentielle de la méthodologie de recherche combinant la découverte causale et le double apprentissage automatique pour l’analyse de la croissance du PIB dans sept pays (2000–2024).

2.3 Prétraitement des Données

2.3.1 Harmonisation des Sources

La première étape du pipeline consiste à harmoniser les trois sources en un dataset unique au format *wide* (une ligne par pays et par année). Cette étape comprend la transformation des fichiers WEO du format large vers le format long, le mapping des codes WEO vers les noms de variables, l’harmonisation des noms de pays entre sources, et la fusion des données WEO et IFS par pays et par année. Le dataset harmonisé final contient **175 observations** (7×25) et **9 colonnes** (Country, Year et

les 7 variables).

2.3.2 Transformations Statistiques

Les transformations appliquées suivent les pratiques standards en macroéconomie empirique. Elles visent à réduire l'hétéroscédasticité et à stabiliser les variances [Maitra \(2025\)](#). Le Tableau 2.2 résume ces choix pour chaque variable :

TABLE 2.2 – Transformations appliquées par variable

Variable	Transformation	Justification
GDP_Growth	Niveau	Déjà en %
Capital_Formation	$\log(x)$	$x > 0$, réduire asymétrie
Inflation	$\log(1 + x)$	Gérer les valeurs nulles
Government_Debt	$\log(x)$	$x > 0$, réduire l'asymétrie
Trade_Balance	Niveau	Valeurs négatives possibles
Exchange_Rate	$\log(x)$	$x > 0$, réduire asymétrie
Reserves	$\log(x)$	$x > 0$, réduire asymétrie

2.3.3 Winsorisation

Après transformation, une winsorisation aux quantiles 1% à 99% est appliquée à chaque variable numérique, séparément par pays. Ce choix est justifié par la forte hétérogénéité des distributions entre pays développés et pays en développement : un outlier pour la France ne revêt pas la même signification économique qu'un outlier pour la RDC.

2.3.4 Standardisation Z-score

La standardisation Z-score est appliquée à l'ensemble des variables numériques après winsorisation. L'Équation (2.1) présente la transformation utilisée :

$$z = \frac{x - \mu}{\sigma} \quad (2.1)$$

Cette transformation garantit que toutes les variables ont une moyenne nulle ($\mu = 0$) et un écart-type unitaire ($\sigma = 1$). La comparabilité des effets estimés par le DML entre pays aux structures économiques très différentes en dépend directement. Le Tableau 2.3 présente les statistiques descriptives globales du dataset standardisé :

TABLE 2.3 – Statistiques descriptives globales (données standardisées, $n = 174$ ²)

Variable	Mean	Std	Min	Max	Skew	Kurtosis
GDP_Growth	0.00	1.00	-2.89	2.65	0.12	0.95
Capital_Formation	0.00	1.00	-3.04	1.48	-1.20	1.22
Inflation	0.00	1.00	-1.63	3.22	0.58	0.39
Government_Debt	0.00	1.00	-2.99	1.75	-0.84	0.50
Trade_Balance	0.00	1.00	-1.86	2.96	0.55	0.24
Exchange_Rate	0.00	1.00	-1.18	1.81	0.28	-1.41
Reserves	0.00	1.00	-2.81	0.99	-1.27	0.61

La Figure 3.3 (présentée en Section 3.2) affiche la matrice de corrélation des variables standardisées. La corrélation quasi nulle entre Capital_Formation et GDP_Growth ($r = -0.02$) justifie le recours au DML pour isoler l'effet causal net.

2.4 Structure Temporelle Causale

2.4.1 Justification du Lagging

Une contribution méthodologique centrale de ce travail est l'imposition d'une structure temporelle causale explicite entre le traitement et l'outcome. Toutes les variables explicatives (traitement et confondeurs) sont prises en année $t - 1$, tandis que l'outcome est mesuré en année t . Les Équations (2.2) et (2.3) formalisent cette structure :

$$\underbrace{\text{Capital_Formation}(t-1)}_{\text{Traitement}} \longrightarrow \underbrace{\text{GDP_Growth}(t)}_{\text{Outcome}} \quad (2.2)$$

$$\underbrace{X(t-1)}_{\text{Confondeurs}} \longrightarrow \underbrace{\text{GDP_Growth}(t)}_{\text{Outcome}} \quad (2.3)$$

Cette structure repose sur trois arguments complémentaires. L'argument de temporalité : le décalage d'un an impose une direction temporelle claire, éliminant la causalité inverse contemporaine [Chernozhukov et al. \(2018\)](#). L'argument économique : la théorie keynésienne de l'accélérateur prédit que les effets de l'investissement sur la production se manifestent avec un délai, les infrastructures acquises en $t - 1$ devenant productives en t [Maitra \(2025\)](#). L'argument statistique : la corrélation contemporaine entre Capital_Formation et GDP_Growth est quasi nulle ($r = -0.02$), suggérant que l'effet causal ne peut être capturé qu'avec un décalage temporel approprié.

2.4.2 Fichiers Produits

Cette transformation génère deux fichiers distincts selon l'usage. Le fichier `data_transformed_unscaled_lagged` contient les données transformées, non standardisées, avec noms de colonnes explicites (`Capital_Formation_lag1`, `Inflation_lag1`, etc.); il est utilisé pour le Causal Discovery (Granger, VAR, VECM). Le fichier `data_prepared_for_dml_lagged.csv` contient les données transformées, standardisées, avec variables laggées renommées vers leurs noms usuels; il est utilisé pour le Double Machine Learning.

2.4.3 Impact sur les Observations

L'application du lag d'une année entraîne la perte de la première observation de chaque pays (année 2000), réduisant le dataset de **175 à 168 observations** théoriques (7×24 , période 2001–2024). Après suppression des valeurs manquantes résiduelles issues des transformations logarithmiques, le dataset final utilisé dans les analyses DML contient **167 observations** effectives. Cette perte d'une observation supplémentaire correspond à une valeur manquante pour la RDC en 2024 après transformation.

2. Le dataset initial contient 175 observations (7×25). Une observation de la RDC a été perdue lors des transformations logarithmiques (valeur manquante), portant n à 174 pour les statistiques descriptives. Après lagging (perte de 2000) et suppression des NAs résiduels, le dataset DML final contient 167 observations.

2.5 Phase 1 : Pipeline de Causal Discovery

2.5.1 Vue d'Ensemble

La Phase 1 vise à identifier les structures causales entre les 7 variables macroéconomiques pour chacun des 7 pays. Elle combine quatre méthodes complémentaires dont les résultats sont ensuite synthétisés en un graphe causal (DAG) de consensus multi-méthodes. Chaque méthode est appliquée séquentiellement et indépendamment par pays.

2.5.2 Étape 1 : Test de Causalité de Granger

Le test de Granger est appliqué à toutes les paires de variables ($7 \times 6 = 42$ paires par pays) sur les données `data_transformed_unscaled_lagged.csv`. Pour chaque paire (X, Y) , la p-value minimale sur les lags 1 à 3 est retenue. Un lien $X \rightarrow Y$ est déclaré significatif si cette p-value minimale est inférieure au seuil $\alpha = 0.05$. Les résultats sont exportés sous forme de matrice de p-values et de liste d'arêtes significatives par pays.

2.5.3 Étape 2 : Modèle VAR

Le modèle VAR est estimé par pays sur les données `data_transformed_unscaled_lagged.csv` avec un lag maximal initial de 4. La sélection du lag optimal procède en trois étapes : estimation des $\text{VAR}(p)$ pour $p = 1, \dots, p_{\max}$ avec $p_{\max} = \min(4, \lfloor (T - 5)/k \rfloor)$ afin d'éviter la surparamétrisation ; sélection du lag minimisant le critère BIC parmi les estimations réussies ; estimation des fonctions de réponse impulsionnelle (IRF) et de la décomposition de la variance (FEVD) sur un horizon de 10 périodes.

Le lag optimal retenu est 2 pour 6 pays sur 7 (France, Allemagne, Angola, Ghana, Maroc, Nigeria) et 1 pour la RDC.

2.5.4 Étape 3 : Test de Johansen et VECM

Avant l'estimation du VECM, le test de cointégration de Johansen est appliqué à chaque pays pour déterminer le rang de cointégration r au seuil de 95%. Le VECM est ensuite estimé avec $k_{ar_diff} = p_{\text{VAR}} - 1$ et le rang r issu du test de Johansen, en utilisant un terme déterministe de type ci (constante dans la relation de cointégration). Les vecteurs α (vitesses d'ajustement) et β (vecteurs de cointégration) sont exportés par pays.

2.5.5 Étape 4 : PCMCI

Contrairement aux trois méthodes précédentes, PCMCI est appliqué sur les données **non laggées** `data_transformed_unscaled.csv` : l'algorithme gère lui-même la structure temporelle des lags [Runge et al. \(2019\)](#). Un fichier `pcmci_<pays>.csv` est préparé par pays avant l'exécution. Les paramètres retenus sont : $\tau_{\min} = 1$, $\tau_{\max} = 3$, $\alpha_{\text{PC}} = 0.05$, test CMiknn avec $k = 5$ voisins et correction FDR Benjamini-Hochberg.

2.5.6 Étape 5 : Synthèse DAG Multi-Méthodes

Les résultats des 4 méthodes sont synthétisés en un DAG final par pays. Pour chaque arête potentielle ($X \rightarrow Y$), un score de support s est calculé comme le nombre de méthodes ayant détecté ce lien de manière significative. L'Équation (2.4) présente le calcul de ce score :

$$s(X \rightarrow Y) = \mathbf{1}_{\text{Granger}} + \mathbf{1}_{\text{VAR}} + \mathbf{1}_{\text{PCMCI}} \quad (2.4)$$

où $\mathbf{1}_{\text{méthode}}$ vaut 1 si la méthode détecte le lien et 0 sinon. Le score varie entre 0 et 3. Le VECM n'est pas intégré dans ce score car il ne produit pas de tests d'arêtes causales dirigées comparables aux trois autres méthodes. Il fournit néanmoins des informations complémentaires sur l'exogénéité des variables via le vecteur α (voir Section 3.3.3). Seules les arêtes avec $s \geq 1$ sont retenues dans le DAG final. Le signe de chaque arête est déterminé par consensus entre les méthodes et codé par une couleur distincte : vert (positif), rouge (négatif), orange (mixte), gris (inconnu).

2.6 Phase 2 : Pipeline Double Machine Learning

2.6.1 Vue d'Ensemble

La Phase 2 vise à estimer l'effet causal quantifié de la Formation de Capital sur la croissance du PIB, en contrôlant rigoureusement les confondeurs macroéconomiques. Elle comprend cinq étapes séquentielles : construction de la matrice DML, évaluation des modèles nuisances, estimation de l'ATE global, estimation des effets par pays et par groupe, et analyse de l'hétérogénéité individuelle via le Causal Forest.

2.6.2 Construction de la Matrice DML

La matrice DML est construite à partir du fichier `data_prepared_for_dml_lagged.csv`. Elle comprend : $Y = \text{GDP_Growth}(t)$ (variable outcome); $T = \text{Capital_Formation}(t-1)$ (variable de traitement); X_{macro} , soit cinq confondeurs en $t - 1$ (Inflation, Government_Debt, Trade_Balance, Exchange_Rate, Reserves); et X_{pays} , soit six variables indicatrices pays obtenues par encodage one-hot de la variable Country, avec l'Allemagne comme catégorie de référence (*drop_first=True*).

La matrice finale contient ainsi **167 observations** et **11 variables** ($5 + 6$) dans X .

Le choix de ces cinq confondeurs macroéconomiques répond à un critère de disponibilité harmonisée sur l'ensemble du panel et sur les 25 années considérées. Il laisse toutefois hors du modèle une catégorie de déterminants potentiellement importants : les facteurs institutionnels (qualité de la gouvernance, contrôle de la corruption, stabilité politique, État de droit). Ces variables influencent plausiblement à la fois le niveau d'investissement et la croissance, et constituent donc des confondeurs non observés au sens de l'hypothèse d'*unconfoundedness* (Section 4). L'inclusion d'effets fixes pays absorbe la composante structurelle et invariante dans le temps de ces facteurs, mais non leur évolution sur la période. Cette limite est discutée en Section 3.4.6 et constitue l'une des principales pistes d'extension de ce travail.

2.6.3 Évaluation des Modèles Nuisances

Avant l'estimation de l'ATE, six modèles candidats sont évalués en validation croisée 5-fold pour les deux équations nuisances $E[Y|X]$ et $E[T|X]$. Le Tableau 2.4 présente les résultats obtenus :

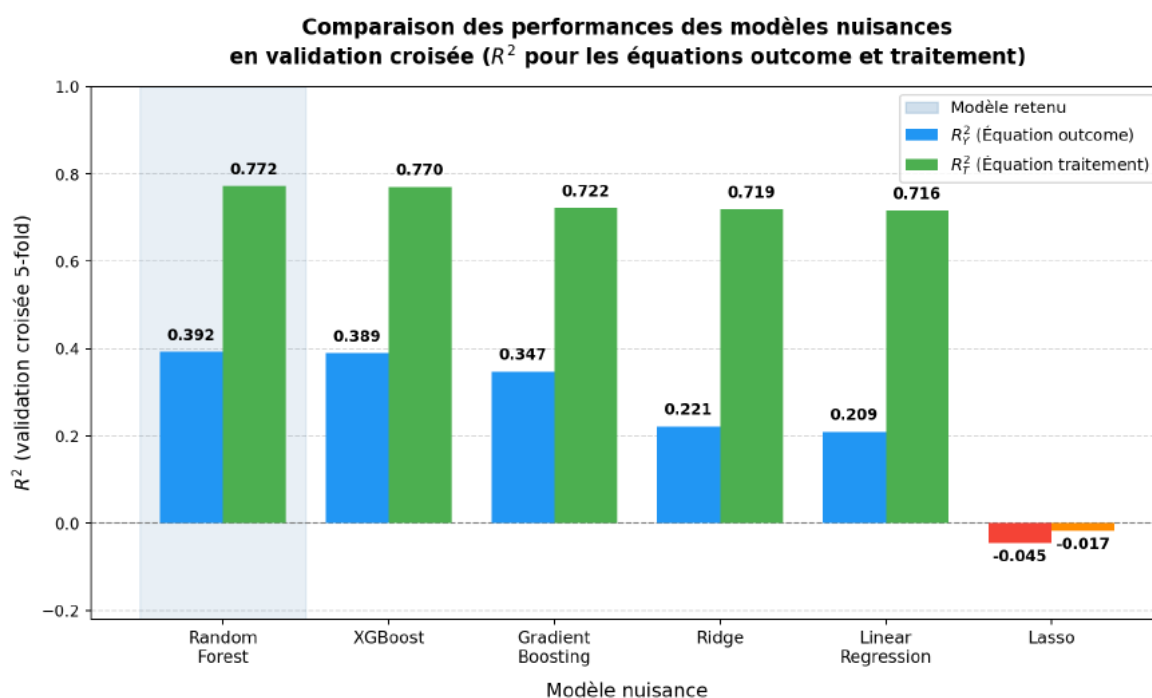
TABLE 2.4 – Performance des modèles nuisances en validation croisée 5-fold

Modèle	R_Y^2	R_T^2	MSE_Y	MSE_T
RandomForest	0.392	0.772	0.576	0.220
XGBoost	0.389	0.770	0.574	0.226
GradientBoosting	0.347	0.722	0.611	0.261
Ridge	0.221	0.719	0.743	0.264
LinearRegression	0.209	0.716	0.754	0.268
Lasso	-0.045	-0.017	1.003	0.991

Le RandomForest est retenu comme modèle nuisance pour les deux équations, avec les hyperparamètres suivants : 400 arbres, profondeur maximale 6, minimum 2 observations par feuille, graine aléatoire 42. Ce choix se justifie par sa performance supérieure ($R_Y^2 = 0.392$, $R_T^2 = 0.772$) et sa robustesse aux non-linéarités, face aux alternatives linéaires (Ridge : $R_Y^2 = 0.221$, Lasso : $R_Y^2 = -0.045$) [Oprescu et al. \(2019\)](#).

Le $R_T^2 = 0.772$ mérite une attention particulière : il confirme que la Formation de Capital est fortement déterminée par les confondeurs. Sans contrôle rigoureux, tout estimateur naïf serait sévèrement biaisé, ce qui justifie le recours au DML.

La Figure 2.2 compare les performances des six modèles candidats en validation croisée.



Note : RandomForest retenu ($R^2_Y = 0.392$, $R^2_T = 0.772$). Le Lasso ($R^2 < 0$) est inadapté aux non-linéarités du panel.

FIGURE 2.2 – Comparaison des performances des modèles nuisances en validation croisée 5-fold (R^2 pour les équations outcome $E[Y|X]$ et traitement $E[T|X]$). Le modèle retenu est encadré en bleu.

Le RandomForest ($R^2_Y = 0.392$) et XGBoost ($R^2_Y = 0.389$) dominent largement les modèles linéaires, confirmant la présence de non-linéarités dans les relations macroéconomiques. Le Lasso ($R^2 = -0.045$) s'avère inadapté à ce contexte.

2.6.4 Estimation de l'ATE Global

L'ATE global est estimé par LinearDML (econml) avec cross-fitting 5-fold sur l'ensemble des 167 observations. L'inférence statistique est conduite via la méthode auto de econml, basée sur une approximation normale asymptotique [Oprescu et al. \(2019\)](#).

2.6.5 Estimation par Pays

Un LinearDML séparé est estimé pour chaque pays sur son sous-échantillon de 24 observations. Le nombre de folds est adapté à la taille du sous-échantillon : $K_{pays} = \min(5, \lfloor n/6 \rfloor)$, avec un minimum de 2 folds. Les dummies pays sont exclues de la matrice X pour les estimations par pays ; seuls les 5 confondeurs macroéconomiques sont retenus.

2.6.6 Estimation par Groupe

L'ATE est estimé séparément pour deux groupes de pays : le groupe des pays développés, comprenant la France et l'Allemagne ($n = 48$), et le groupe des pays en développement, comprenant l'Angola,

le Ghana, le Nigeria, le Maroc et la RDC ($n = 119$). L'inférence est conduite via `BootstrapInference` avec 200 échantillons bootstrap [Oprescu et al. \(2019\)](#).

2.6.7 Hétérogénéité Individuelle : CausalForestDML

Les effets de traitement hétérogènes individuels $\tau(x_i)$ sont estimés par CausalForestDML avec les paramètres suivants : 800 arbres, minimum 5 observations par feuille, modèles nuisances RandomForest (500 arbres, profondeur maximale 8), cross-fitting 5-fold [Oprescu et al. \(2019\)](#). La différence entre groupes est ensuite testée par un test de Welch (variances inégales) sur les distributions des effets individuels.

2.7 Conclusion partielle

Ce chapitre a décrit l'ensemble du pipeline méthodologique mis en œuvre dans ce travail. Depuis l'harmonisation des sources brutes (FMI WEO, FMI IFS, Banque Mondiale) jusqu'à l'estimation des effets causaux hétérogènes par CausalForestDML, chaque étape a été conçue pour garantir la rigueur causale et la reproductibilité des résultats. Quatre choix méthodologiques méritent d'être soulignés.

La structure temporelle causale (variables en $t - 1$, outcome en t) constitue le premier garde-fou contre la causalité inverse et la simultanéité contemporaine, deux biais majeurs en macroéconomie observationnelle. La combinaison de quatre méthodes de Causal Discovery en un DAG de consensus multi-méthodes permet de distinguer les liens causaux robustes (confirmés par plusieurs méthodes, $s \geq 2$) des liens fragiles (détectés par une seule méthode, $s = 1$), réduisant le risque de faux positifs. Le choix du RandomForest comme modèle nuisance est justifié empiriquement par sa performance supérieure ($R_Y^2 = 0.392$, $R_T^2 = 0.772$) face aux alternatives linéaires, confirmant la présence de non-linéarités importantes dans les relations macroéconomiques du panel. L'estimation séparée par pays, par groupe et par individu (CausalForest) permet de capturer l'hétérogénéité structurelle entre les 7 économies étudiées, évitant le piège du modèle universel (*one-size-fits-all*), que la littérature a montré inadapté au contexte macroéconomique africain [Mulomba et al. \(2025\)](#); [Biaba et al. \(2025\)](#).

Le chapitre suivant présente et analyse l'ensemble des résultats obtenus, en commençant par les structures causales identifiées (Phase 1) puis les effets causaux estimés (Phase 2).

RÉSULTATS ET DISCUSSION

3.1 Introduction

Ce chapitre couvre deux phases d'analyse. La Phase 1 documente les résultats du Causal Discovery : tests de Granger, modèles VAR, VECM, algorithme PCMCi et synthèse DAG. La Phase 2 rapporte les résultats du Double Machine Learning : ATE global, ATE par pays, ATE par groupe et effets hétérogènes individuels. Chaque résultat est accompagné d'une interprétation économique ancrée dans les théories macroéconomiques établies.

3.2 Analyse Descriptive des Données

3.2.1 Évolution Temporelle des Variables

L'évolution temporelle des variables sur la période 2000–2024 fait apparaître des profils fortement contrastés. Les trajectoires des pays développés diffèrent structurellement de celles des pays en développement.

La Figure 3.1 illustre l'évolution du taux de croissance annuel du PIB réel pour les 7 pays sur la période 2000–2024.

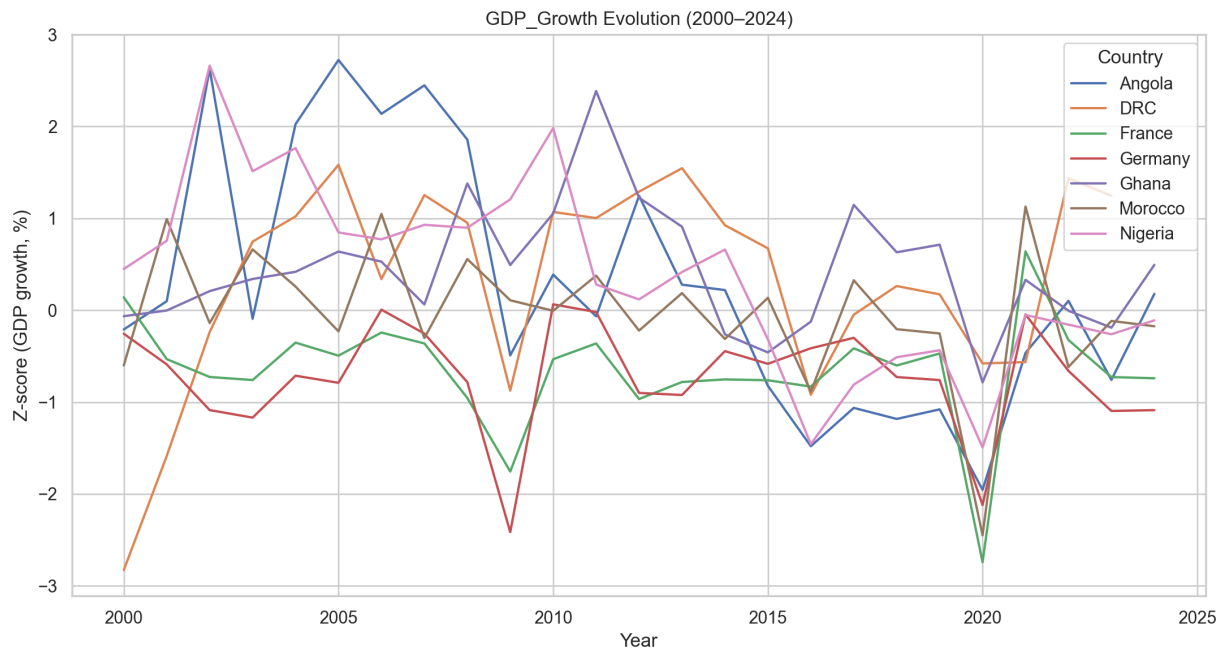


FIGURE 3.1 – Évolution du taux de croissance annuel du PIB réel par pays (2000–2024, valeurs standardisées, Z-score). Source : FMI WEO et Banque Mondiale.

Trois faits stylisés ressortent de la Figure 3.1, dont les valeurs sont exprimées en écarts-types par rapport à la moyenne du panel. La crise financière mondiale de 2008–2009 provoque un effondrement marqué en Allemagne et en France, tandis que les pays africains affichent une résistance relative aux turbulences des marchés financiers internationaux. Le choc COVID-19 de 2020 est simultané et sévère dans l'ensemble du panel : la France et l'Allemagne y enregistrent leurs valeurs les plus basses de la période, avec une reprise rapide dès 2021 dans la plupart des pays. Sur le long terme, les pays africains (Angola, Ghana, Nigeria, RDC) affichent structurellement des taux de croissance plus élevés que les pays européens, avec une volatilité nettement supérieure : l'Angola enregistre la croissance la plus forte du panel lors du boom pétrolier de 2002, mais aussi les baisses les plus prononcées lors des chocs sur les matières premières.

La Figure 3.2 illustre l'évolution de la Formation Brute de Capital pour les 7 pays sur la période 2000–2024.

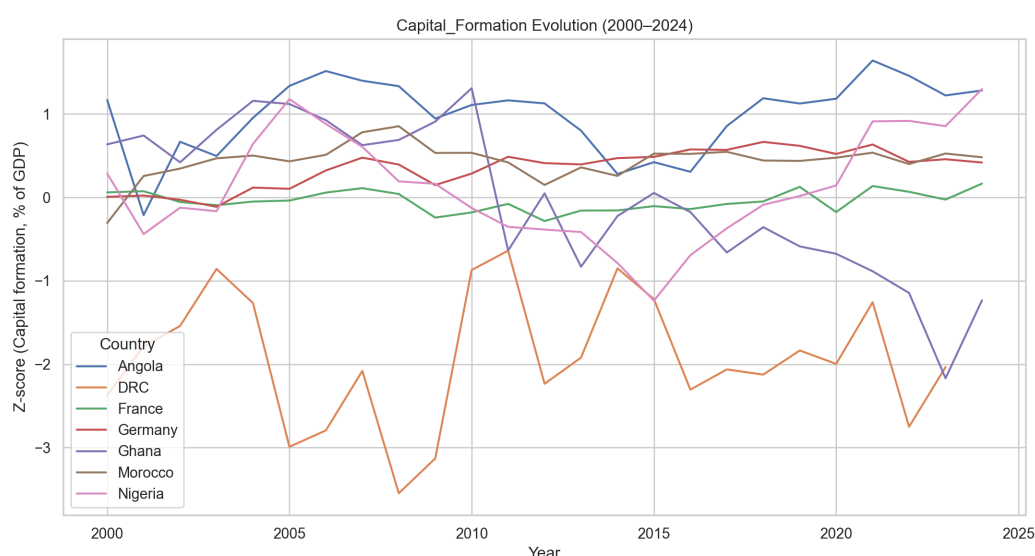


FIGURE 3.2 – Évolution de la Formation Brute de Capital par pays (2000–2024, valeurs standardisées, Z-score). Source : FMI WEO et Banque Mondiale.

La Figure 3.2 met en évidence une hétérogénéité structurelle marquée dans les niveaux d'investissement. La RDC enregistre les valeurs les plus basses du panel, systématiquement inférieures à la moyenne d'environ trois écarts-types, assorties d'une volatilité extrême. Cette instabilité traduit la fragilité économique et institutionnelle chronique de ce pays. À l'opposé, l'Angola maintient les niveaux les plus élevés du panel, portés par les investissements massifs dans le secteur pétrolier. La France et l'Allemagne se distinguent par des trajectoires stables et convergentes, proches de la moyenne du panel.

3.2.2 Matrice de Corrélation

La Figure 3.3 présente la matrice de corrélation des 7 variables standardisées ($n = 174$).

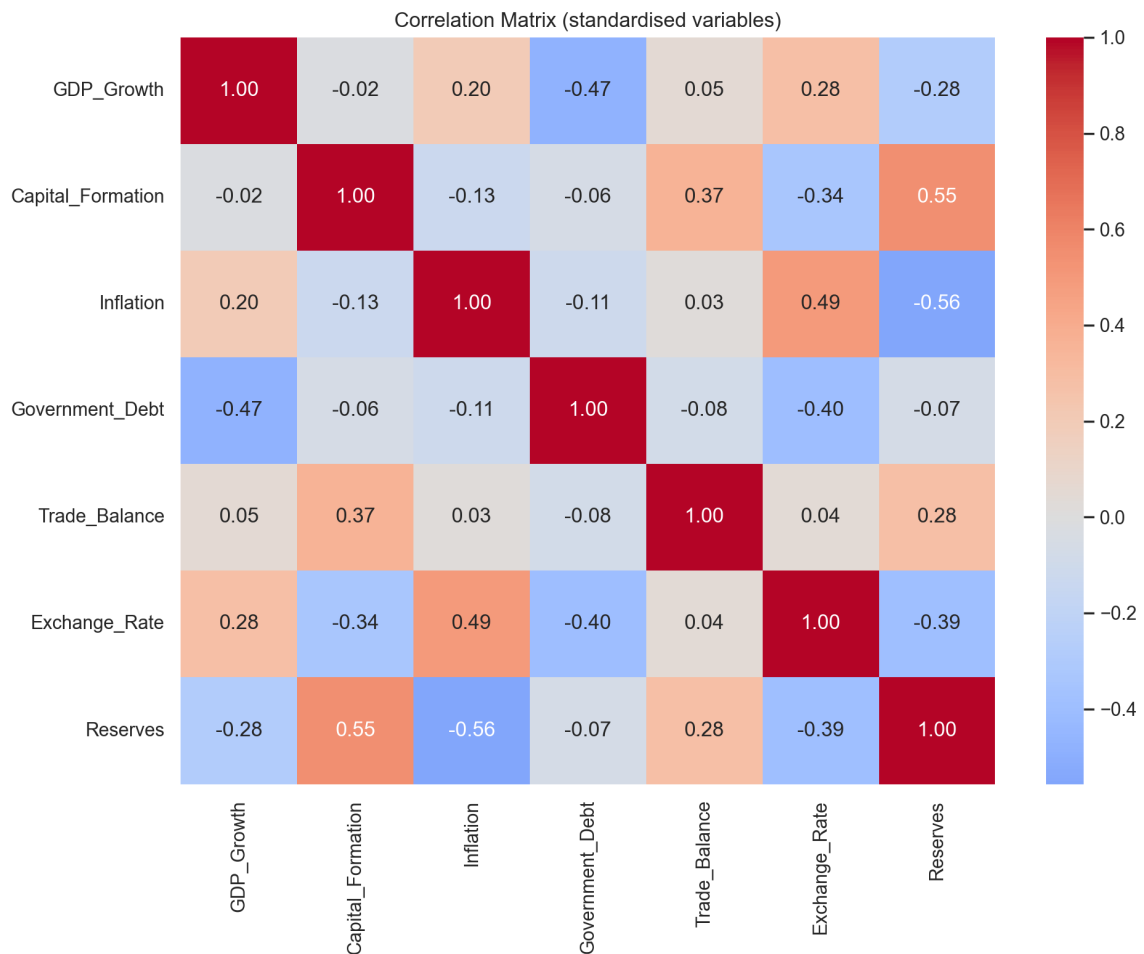


FIGURE 3.3 – Matrice de corrélation des 7 variables standardisées ($n = 174$).

La Figure 3.3 révèle plusieurs associations importantes entre variables. La corrélation entre Capital_Formation et GDP_Growth est quasi nulle ($r = -0.02$) : cette absence de corrélation contemporaine confirme que l'effet causal de l'investissement sur la croissance ne peut être capturé par une simple régression linéaire, et justifie le recours au DML avec structure temporelle laggée. La corrélation entre Government_Debt et GDP_Growth est négativement forte ($r = -0.47$) : une dette publique élevée est associée à une croissance plus faible, cohérent avec la littérature sur le seuil de soutenabilité [Mazumder et al. \(2025\)](#). La corrélation entre Capital_Formation et Reserves est positivement forte ($r = +0.55$) : les pays qui investissent davantage tendent à accumuler plus de réserves de change. La corrélation entre Inflation et Reserves est négativement forte ($r = -0.56$) : les pays à forte inflation disposent généralement de réserves plus faibles. Enfin, la corrélation entre Exchange_Rate et Inflation est positive modérée ($r = +0.49$), reflétant l'effet pass-through dans les économies africaines très ouvertes.

3.2.3 Relation Capital Formation et GDP Growth

La Figure 3.4 présente le nuage de points entre Capital_Formation et GDP_Growth en valeurs standardisées (Z-score).

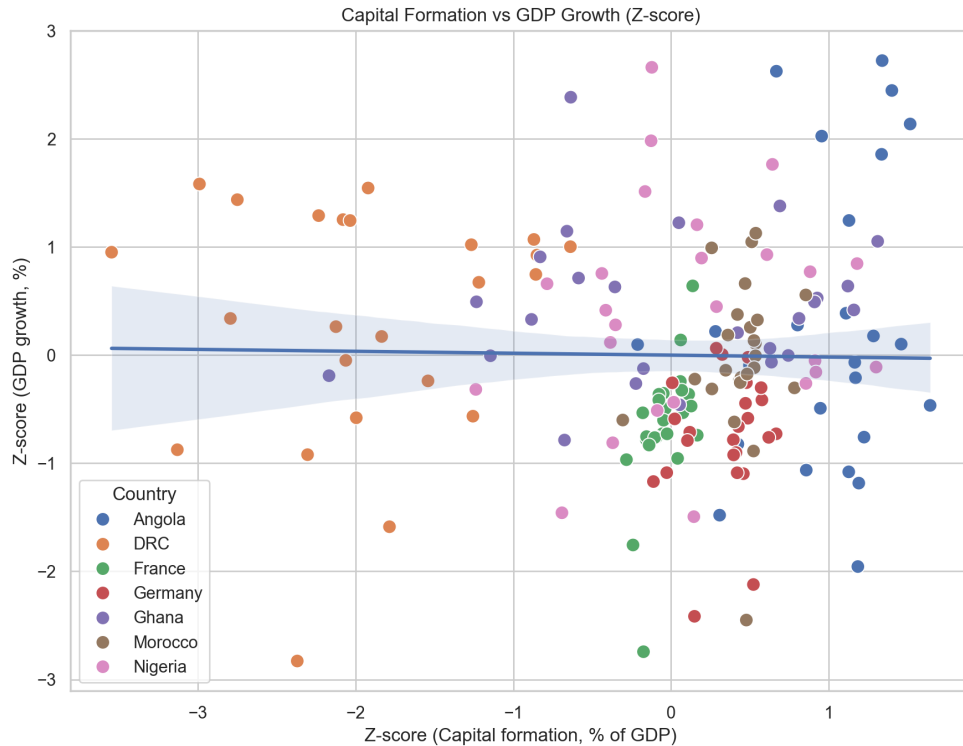


FIGURE 3.4 – Nuage de points Capital_Formation vs GDP_Growth (valeurs standardisées, Z-score).

La droite de tendance globale est quasi horizontale ($r = -0.02$), confirmant l’absence de relation linéaire contemporaine entre les deux variables.

La Figure 3.4 confirme visuellement ce constat. La dispersion très large des points révèle une hétérogénéité marquée entre pays. La tendance globale quasi plate traduit l’incapacité d’une simple régression linéaire à saisir l’effet de l’investissement sur la croissance. Une approche causale conditionnelle est indispensable pour isoler cet effet net des influences confondantes.

3.3 Résultats de la Phase 1 : Causal Discovery

3.3.1 Résultats du Test de Granger

3.3.1.1 Structures Causales Identifiées

Le test de Granger (lags 1 à 3, $\alpha = 0.05$) révèle des structures causales contrastées entre pays. Le Tableau 3.1 présente le nombre d’arêtes significatives et la p-value du lien Capital_Formation_lag1 $\rightarrow$ GDP_Growth pour chacun d’eux :

TABLE 3.1 – Résultats du test de Granger par pays (lag 1–3, $\alpha = 0.05$)

Pays	Arêtes sig.	p-val CF→PIB	Significatif
Angola	24	0.188	Non
Maroc	11	0.507	Non
Nigeria	11	0.185	Non
France	9	0.475	Non
Ghana	9	0.097	Non
Allemagne	8	0.352	Non
RDC	10	0.619	Non

Le résultat le plus saillant est l’absence de causalité de Granger entre Capital_Formation_lag1 et GDP_Growth dans les 7 pays du panel. Ce constat s’inscrit dans la continuité de la corrélation contemporaine quasi nulle ($r = -0.02$) documentée à l’étape descriptive.

La Figure 3.5 présente la heatmap des p-values Granger minimales (lags 1–3) pour l’Allemagne.

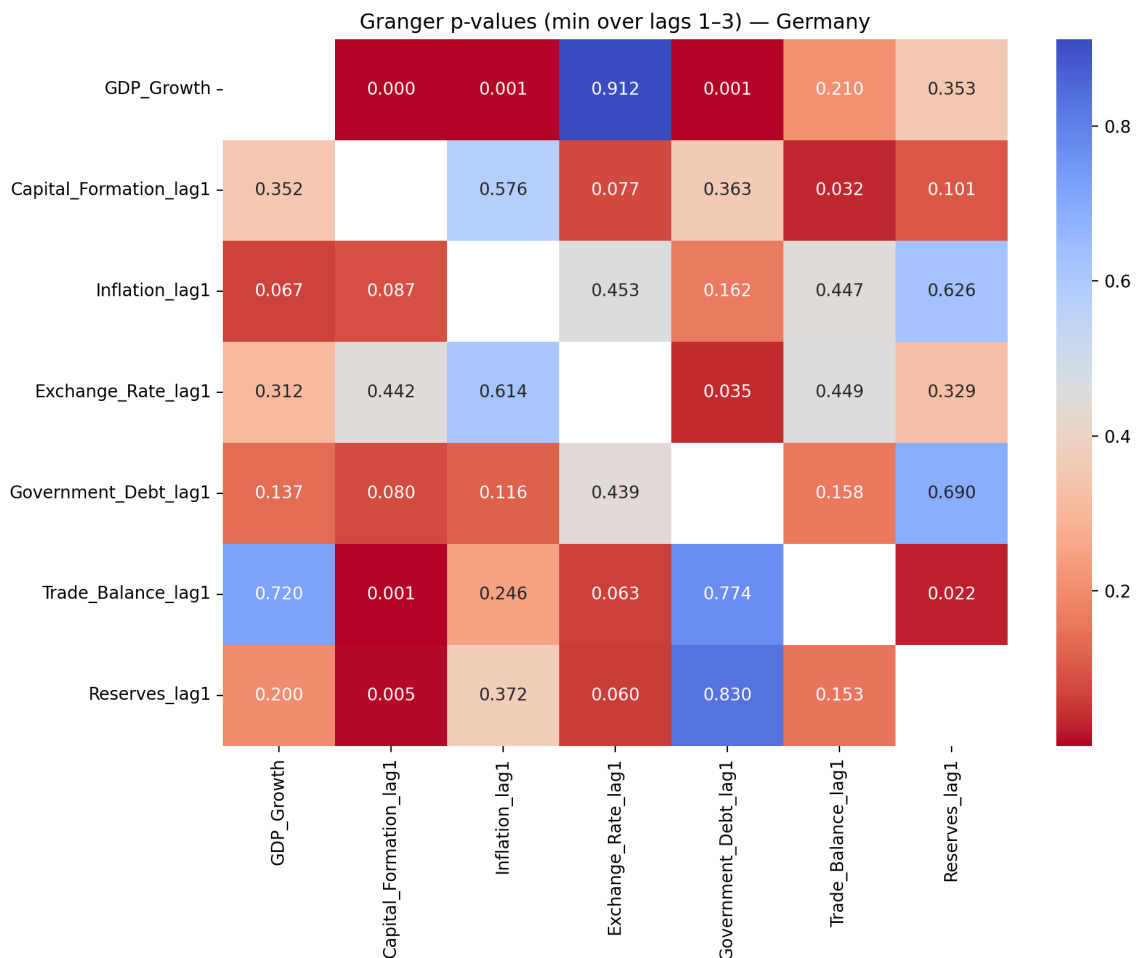


FIGURE 3.5 – Heatmap des p-values Granger minimales (lags 1–3) Allemagne.

Les cellules rouges signalent des liens significatifs ($p < 0.05$). La cellule Capital_Formation_lag1 → GDP_Growth affiche $p = 0.352$, valeur non significative.

3.3.1.2 Interprétation Économique

Plusieurs mécanismes expliquent l'absence de causalité de Granger entre investissement et croissance dans les 7 pays. Le test de Granger est linéaire et bivarié : il ne tient pas compte des confondeurs macroéconomiques qui médiatisent la relation entre investissement et croissance. Par ailleurs, dans les économies africaines, l'investissement formel enregistré par les statistiques officielles ne représente qu'une fraction du total réel. Une large part s'effectue dans le secteur informel, hors du champ de mesure [Mulomba et al. \(2025\)](#). Le canal le plus fréquemment identifié par Granger est le taux de change, présent dans 5 pays sur 7. Ce résultat confirme la contrainte extérieure comme facteur limitant principal des économies africaines.

3.3.2 Résultats du Modèle VAR

3.3.2.1 Sélection des Lags

Le Tableau 3.2 indique le lag optimal retenu par critère BIC pour chacun des 7 pays :

TABLE 3.2 – Lag optimal sélectionné par critère BIC pour chaque pays

Pays	Lag BIC
Allemagne	2
Angola	2
France	2
Ghana	2
Maroc	2
Nigeria	2
RDC	1

Le lag 2 est optimal pour 6 pays sur 7. La RDC retient un lag de 1. Ce résultat reflète la faiblesse des structures dynamiques dans ses séries macroéconomiques, vraisemblablement liée à la dominance des chocs exogènes (conflits, chocs miniers) sur les mécanismes de transmission internes [Mulomba et al. \(2025\)](#).

3.3.2.2 Décomposition de la Variance (FEVD)

La FEVD de GDP_Growth montre que cette variable s'explique elle-même à hauteur d'environ 74% de sa variance à l'horizon 1. Cette forte inertie traduit la persistance de la croissance économique à court terme. La contribution de Capital_Formation_lag1 à la variance de GDP_Growth demeure faible, entre 3 et 4% aux horizons courts. Ce résultat va dans le même sens que les conclusions du test de Granger.

La Figure 3.6 présente la décomposition de la variance de l'erreur de prévision (FEVD) de l'Allemagne pour un lag de 2 et un horizon de 10 périodes.

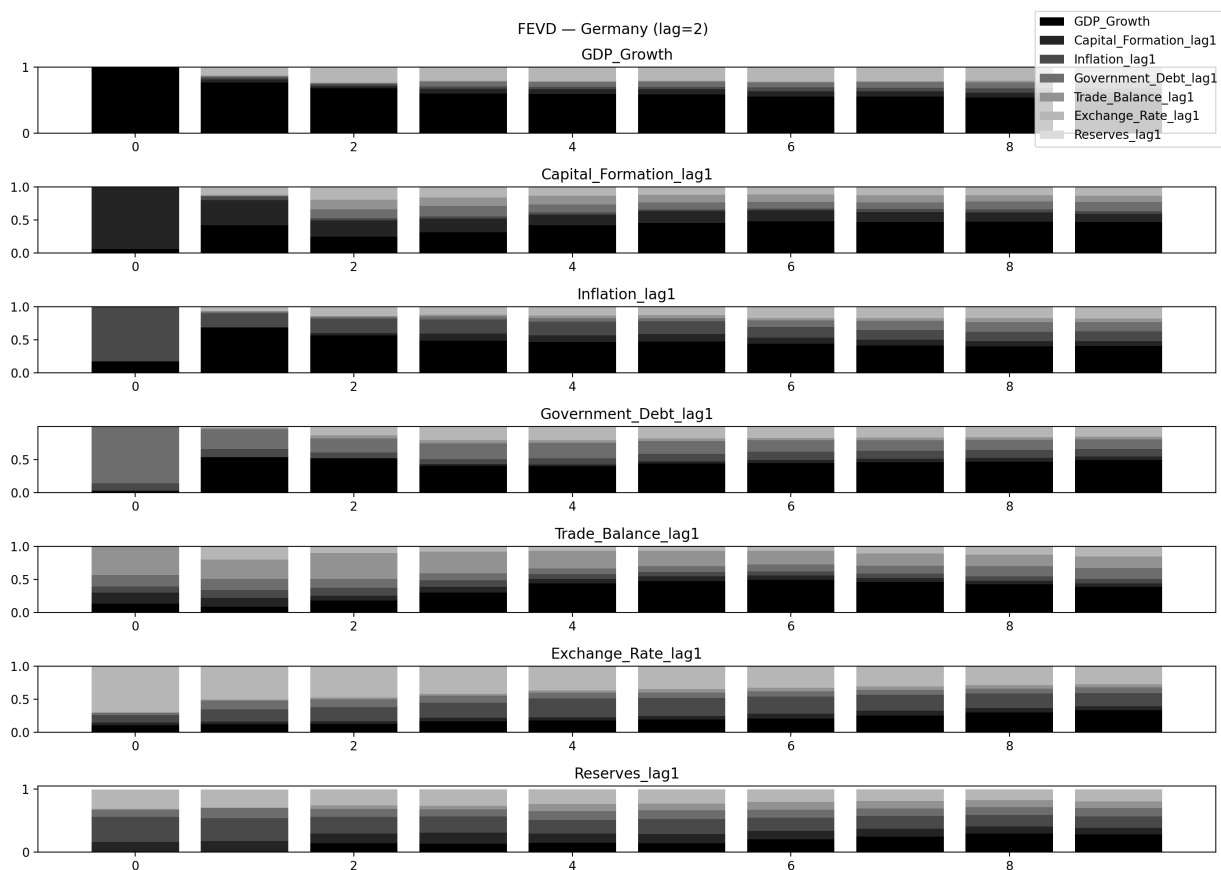


FIGURE 3.6 – Décomposition de la variance de l’erreur de prévision (FEVD) de l’Allemagne (lag=2, horizon=10).

La barre noire représente la part auto-explicée de chaque variable. La contribution de Capital_Formation à GDP_Growth reste inférieure à 4%.

3.3.3 Résultats du Test de Johansen et du VECM

3.3.3.1 Test de Cointégration

Le test de Johansen est ici appliqué à la France à titre illustratif. Il révèle un rang de cointégration plein ($r = k - 1 = 6$, pour $k = 7$ variables). Ce résultat signifie que les 7 variables du système partagent des relations d’équilibre de long terme robustes. Le rang maximal théorique dans un système de k variables est $k - 1$: un rang plein indique que toutes les séries participent à au moins une relation de cointégration. Le Tableau 3.3 présente les statistiques de trace et les valeurs critiques à 95% :

TABLE 3.3 – Test de Johansen : statistiques de trace et valeurs critiques à 95% — France (rang plein $r = k - 1 = 6$, avec $k = 7$ variables)

H_0 (rang $\leq$)	Trace	CV 95%	Résultat
0	291.11	125.62	Rejeté
1	194.91	95.75	Rejeté
2	122.10	69.82	Rejeté
3	65.89	47.85	Rejeté
4	38.15	29.80	Rejeté
5	16.44	15.49	Rejeté
6	6.39	3.84	Rejeté

3.3.3.2 Vecteurs d'Ajustement et de Cointégration

L'estimation du VECM produit les vecteurs α (vitesses d'ajustement) et β (relations de long terme). La relation de cointégration normalisée s'écrit :

$$\begin{aligned}
 & \text{GDP_Growth} + 6.39 \text{ Capital_Formation} \\
 & \quad - 6.69 \text{ Inflation} \\
 & \quad + 8.67 \text{ Government_Debt} \\
 & \quad - 0.34 \text{ Exchange_Rate} \\
 & \quad - 0.54 \text{ Trade_Balance} \\
 & \quad + 2.19 \text{ Reserves} = 0
 \end{aligned} \tag{3.1}$$

Le vecteur d'ajustement α révèle un résultat déterminant pour notre analyse. Pour GDP_Growth, $\alpha = +0.327$: la vitesse d'ajustement est significative et le PIB s'ajuste rapidement vers l'équilibre de long terme. Pour Capital_Formation, $\alpha \approx 0.0003$, quasi nul : l'investissement ne corrige pas les déséquilibres du système. Il est exogène dans la relation de cointégration, ce qui confirme sa validité comme variable de traitement dans le DML.

3.3.3.3 Interprétation Économique

La cointégration de rang plein ($r = k - 1 = 6$) confirme que les 7 variables macroéconomiques partagent des équilibres de long terme robustes. Le coefficient élevé de Capital_Formation dans le vecteur β ($\beta = 6.39$) indique que l'investissement participe significativement à cet équilibre. C'est néanmoins GDP_Growth qui s'ajuste ($\alpha = 0.327$), et non l'investissement ($\alpha \approx 0.0003$). Ce résultat est cohérent avec la théorie keynésienne de l'accélérateur : la croissance réagit aux déséquilibres, l'investissement non [Maitra \(2025\)](#).

3.3.4 Résultats du PCMCI

3.3.4.1 Synthèse par Pays

L'algorithme PCMCI avec test CMiknn et correction FDR Benjamini-Hochberg produit des graphes causaux temporels distincts pour chaque pays. Le Tableau [3.4](#) résume ces résultats :

TABLE 3.4 – Résultats PCMCI par pays ($\tau_{max} = 3$, $\alpha = 0.05$, CMlknn, FDR BH)

Pays	Liens sig.	CF→PIB
Ghana	34	Absent
Nigeria	8	Absent
France	3	Absent
RDC	3	Absent
Allemagne	0	Absent
Angola	0	Absent
Maroc	0	Absent

Le lien Capital_Formation $\rightarrow$ GDP_Growth est **absent dans les 7 pays** selon PCMCI, confirmant les résultats de Granger. Ce résultat est d'autant plus robuste qu'il est obtenu avec un test non linéaire conditionnel (CMlknn), plus exigeant que le test de Granger linéaire bivarié.

La Figure 3.7 présente le graphe causal temporel obtenu par PCMCI pour le Nigeria, avec 8 liens significatifs.

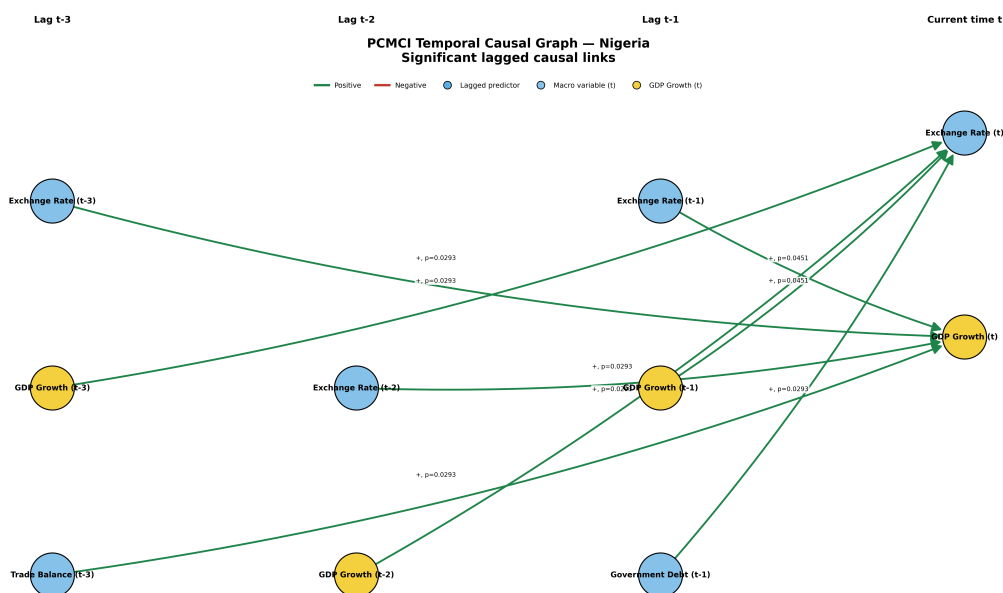


FIGURE 3.7 – Graphe causal temporel PCMCI : Nigeria (8 liens significatifs).

Les flèches vertes indiquent des liens positifs. Le lien dominant est Exchange_Rate $\rightarrow$ GDP_Growth aux lags 1, 2 et 3. Le lien Capital_Formation $\rightarrow$ GDP_Growth est absent.

La Figure 3.8 présente le graphe causal temporel obtenu par PCMCI pour le Ghana.

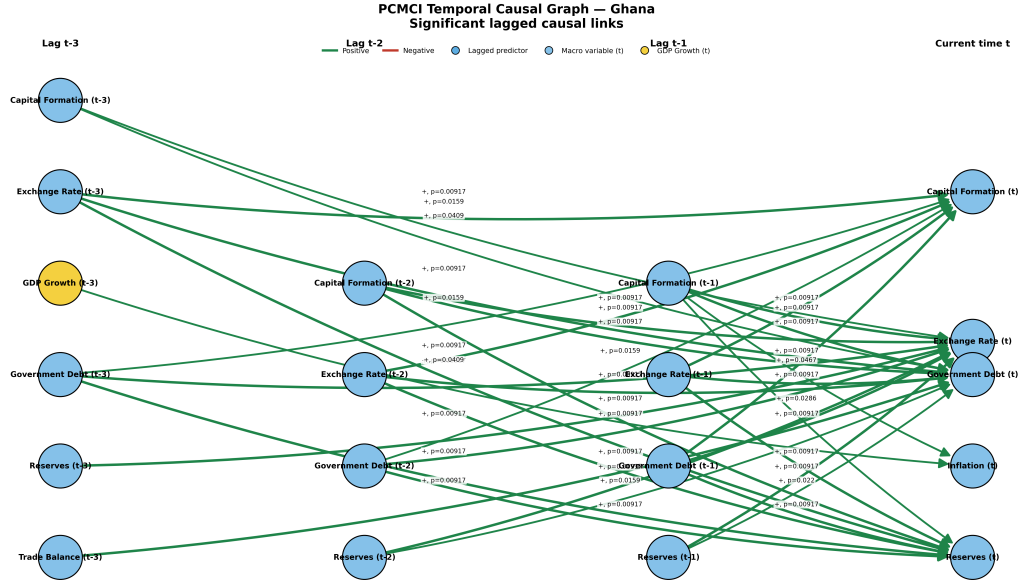


FIGURE 3.8 – Graphe causal temporel PCMCI : Ghana.

Capital_Formation apparaît comme variable source vers d'autres variables (Exchange_Rate, Government_Debt), sans lien direct avec GDP_Growth.

3.3.4.2 Interprétation Économique

Deux enseignements économiques ressortent des résultats PCMCI. Le taux de change apparaît comme la variable causale la plus fréquemment retenue, un résultat repris et discuté dans la synthèse multi-méthodes (Section 3.3.5). L'investissement cause d'autres variables mais non le PIB directement : au Ghana, Capital_Formation est une source active vers Exchange_Rate, Government_Debt et Inflation, et son effet sur la croissance est donc indirect, médiatisé par ces canaux. Les économies les plus stables (Allemagne, Angola, Maroc) n'affichent aucun lien significatif, ce qui peut s'expliquer soit par une puissance statistique limitée du test CMlknn sur $T = 25$ observations, soit par une dominance des chocs exogènes sur les mécanismes de transmission internes.

3.3.5 Synthèse DAG Multi-Méthodes

3.3.5.1 Graphes Causaux Finaux

La synthèse des 4 méthodes produit un DAG final par pays, avec un score de support $s \in \{1, 2, 3\}$. Les Figures 3.9 à 3.11 présentent les DAGs pour les pays les plus représentatifs :

Final DAG — France (support ≥ 1 , acyclic)

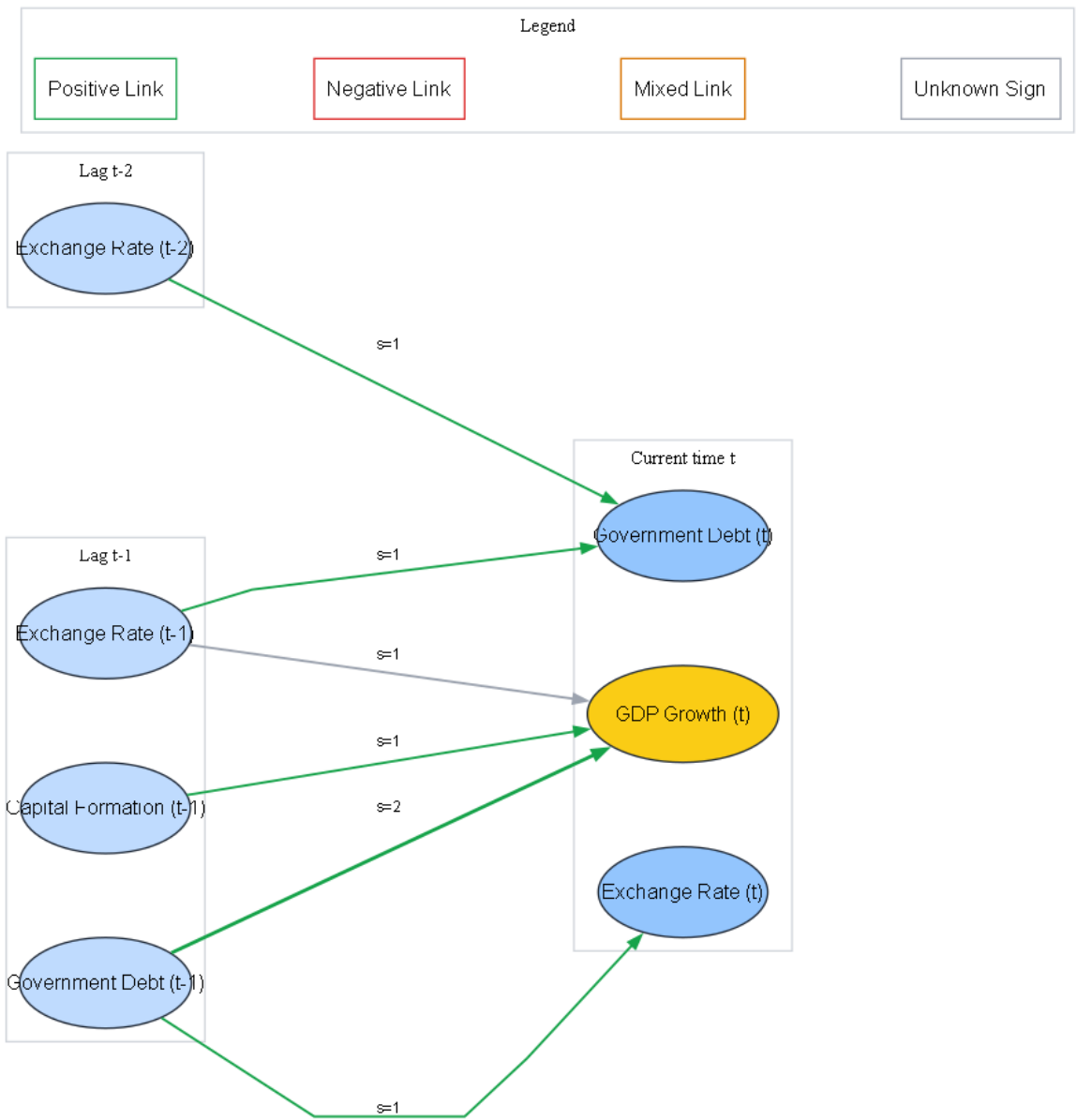


FIGURE 3.9 – DAG final de la France (support ≥ 1 , acyclique).

Capital_Formation(t-1) cause GDP_Growth(t) avec $s = 1$ (signe positif). Government_Debt(t-1) constitue le lien le plus robuste vers GDP_Growth(t) avec $s = 2$.

Final DAG — Angola (support ≥ 1 , acyclic)

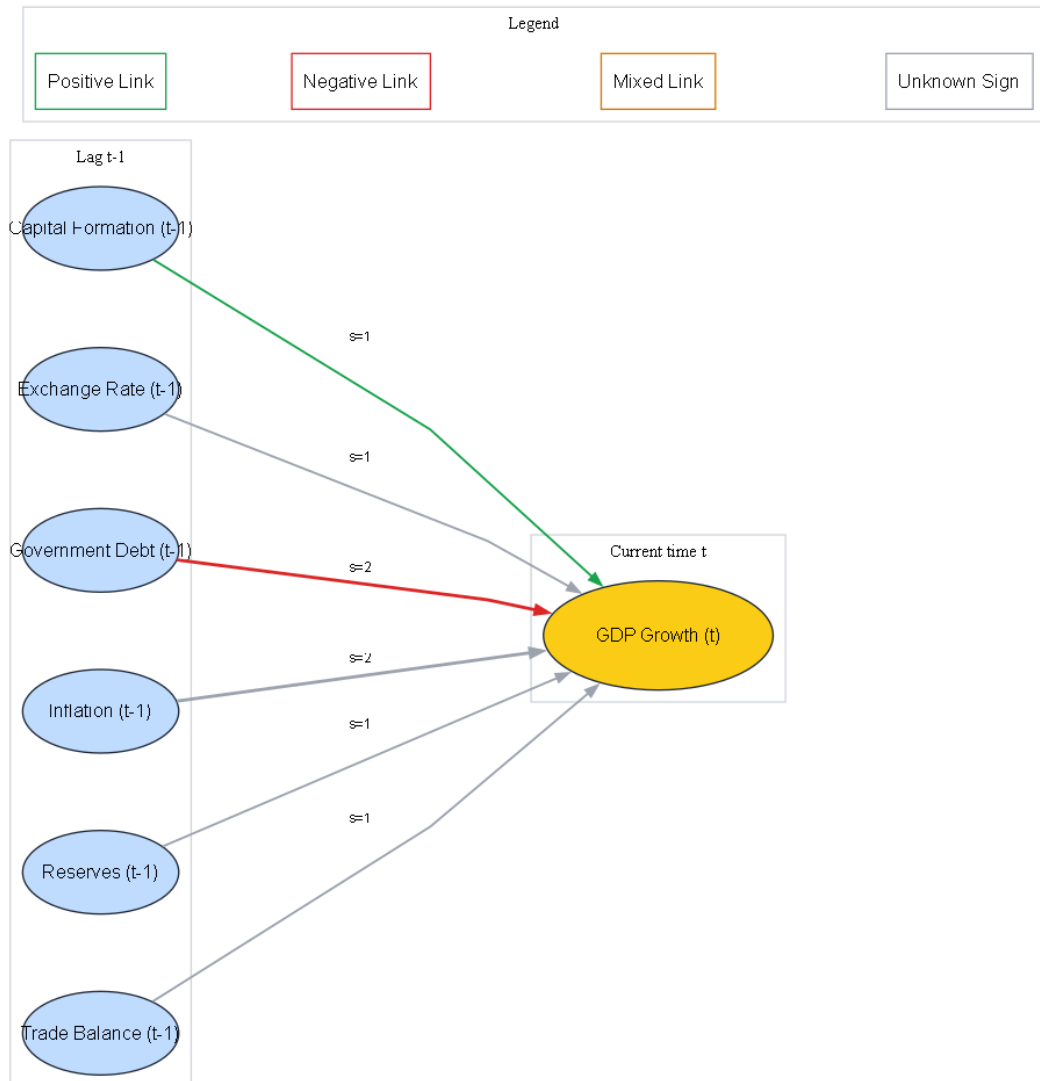


FIGURE 3.10 – DAG final de l'Angola (support ≥ 1 , acyclique).

Capital_Formation(t-1) cause GDP_Growth(t) avec $s = 1$ (signe positif). Government_Debt(t-1) présente l'effet le plus robuste ($s = 2$, signe négatif).

Final DAG — Nigeria (support ≥ 1 , acyclic)



FIGURE 3.11 – DAG final du Nigeria (support ≥ 1 , acyclique).

Exchange_Rate($t-1$) constitue le lien le plus robuste vers GDP_Growth(t) ($s = 2$, signe positif), confirmé conjointement par le test de Granger et par PCMCI. Capital_Formation($t-1$) n'est retenue qu'avec un support minimal ($s = 1$, arête orange), le VAR étant la seule méthode à la détecter et le signe de son coefficient variant selon le lag considéré.

3.3.5.2 Tableau Comparatif Multi-Pays

Le Tableau 3.5 synthétise les résultats des DAGs finaux pour les 7 pays :

TABLE 3.5 – Synthèse des DAGs finaux pour les 7 pays (Causal Discovery multi-méthodes)

Pays	CF→PIB	Support	Signe	Cause principale PIB
Allemagne	Absent	–	–	Gov_Debt(+), Reserves(+)
France	Présent	$s = 1$	+	Gov_Debt(+, $s = 2$)
Maroc	Présent	$s = 1$	?	Gov_Debt(–, $s = 2$)
Ghana	Présent	$s = 1$	+	ExchRate(+), Gov_Debt(+)
Nigeria	Présent	$s = 1$	±	ExchRate(+, $s = 2$), PIB auto
Angola	Présent	$s = 1$	+	Gov_Debt(–, $s = 2$)
RDC	Absent	–	–	ExchRate(+), Gov_Debt(+)

3.3.5.3 Interprétation Économique de la Synthèse DAG

Quatre conclusions ressortent de la synthèse multi-méthodes.

D'abord, Capital_Formation n'est pas un hub causal universel. Le lien CF → PIB est absent dans 2 pays (Allemagne, RDC) et n'est retenu qu'avec un support minimal ($s = 1$) dans les 5 autres (France, Maroc, Ghana, Nigeria, Angola). Ce support provient exclusivement du modèle VAR : ni le test de Granger (Tableau 3.1) ni PCMCI (Tableau 3.4) ne détectent ce lien dans aucun pays du panel. Au Nigeria, le signe demeure ambigu, le coefficient VAR changeant de sens selon le lag considéré. La causalité directe de l'investissement sur la croissance apparaît donc fragile et dépendante du contexte.

Ensuite, Government_Debt est la variable causale la plus robuste du panel : elle cause GDP_Growth dans 6 pays sur 7 (à l'exception du Nigeria) et constitue l'arête de support le plus élevé ($s = 2$) en France, au Maroc et en Angola. Son signe n'est toutefois pas uniforme : positif en France, en Allemagne, au Ghana et en RDC, négatif au Maroc et en Angola. Cette inversion est cohérente avec la thèse d'un seuil de soutenabilité au-delà duquel la relation entre dette et croissance change de sens, seuil dont la position dépend de la structure de financement propre à chaque économie Mazumder et al. (2025).

Troisièmement, Exchange_Rate joue un rôle pivot dans les économies africaines : il cause directement GDP_Growth au Ghana, au Nigeria et en RDC, et constitue par ailleurs le canal le plus fréquemment identifié par le test de Granger, dans 5 pays sur 7. Ce résultat confirme le poids de la contrainte extérieure comme mécanisme de transmission macroéconomique dans ces économies Mulomba et al. (2025).

Enfin, le Causal Discovery ne suffit pas à quantifier l'effet. Même lorsqu'un lien CF → PIB est détecté, il repose sur une seule méthode, son signe reste parfois ambigu et sa magnitude n'est pas estimée. Le recours au Double Machine Learning est donc nécessaire pour passer de l'identification structurelle à la quantification de l'effet net.

3.4 Résultats de la Phase 2 : Double Machine Learning

3.4.1 ATE Global

Le Tableau 3.6 présente les résultats de l'estimation de l'ATE global par LinearDML :

TABLE 3.6 – ATE global avec LinearDML (167 observations, 7 pays, 2000-2024)

Paramètre	Valeur
ATE	−0.528
IC 95% borne inférieure	−1.042
IC 95% borne supérieure	−0.014
Écart-type	0.262
p-value	0.044
Observations	167
Variables dans X	11
Folds CV	5

L'ATE global est **négatif et statistiquement significatif** ($\hat{\theta} = -0.528$, $p = 0.044$). En moyenne sur les 7 pays, une hausse d'un écart-type de la Formation de Capital en $t - 1$ est associée à une réduction de 0.528 point de la croissance du PIB en t , toutes choses égales par ailleurs.

Interprétation économique : Ce résultat négatif peut paraître contre-intuitif. Il s'explique par la forte hétérogénéité structurelle entre pays développés et pays en développement dans le panel. L'effet très négatif de la France (−3.73) domine l'ATE moyen. Ce phénomène illustre le piège du modèle universel (*one-size-fits-all*) : une moyenne sur panel hétérogène peut masquer des effets structurellement opposés selon le contexte économique de chaque pays [Mulomba et al. \(2025\)](#); [Biaba et al. \(2025\)](#).

3.4.2 ATE par Pays

Le Tableau 3.7 présente les ATE estimés par LinearDML pour chacun des 7 pays :

TABLE 3.7 – ATE par pays avec LinearDML

Pays	ATE	IC 95%	StdErr	p-val	Sig.
France	−3.728	[−6.952 ; −0.503]	1.645	0.023	Oui
RDC	+0.542	[+0.078 ; +1.006]	0.237	0.022	Oui
Maroc	+1.704	[−2.538 ; +5.946]	2.164	0.431	Non
Ghana	+0.300	[−0.308 ; +0.909]	0.310	0.334	Non
Allemagne	+0.414	[−1.282 ; +2.111]	0.866	0.632	Non
Angola	+0.113	[−1.963 ; +2.189]	1.059	0.915	Non
Nigeria	+0.008	[−0.520 ; +0.535]	0.269	0.978	Non

L'ATE par pays est estimé sur 24 observations par pays, à l'exception de la RDC qui en compte 23. Le paramètre n_splits est fixé à 3 pour la RDC et à 4 pour les autres pays.

La Figure 3.12 présente les effets causaux moyens (ATE) estimés par LinearDML pour chaque pays, avec leurs intervalles de confiance à 95%.

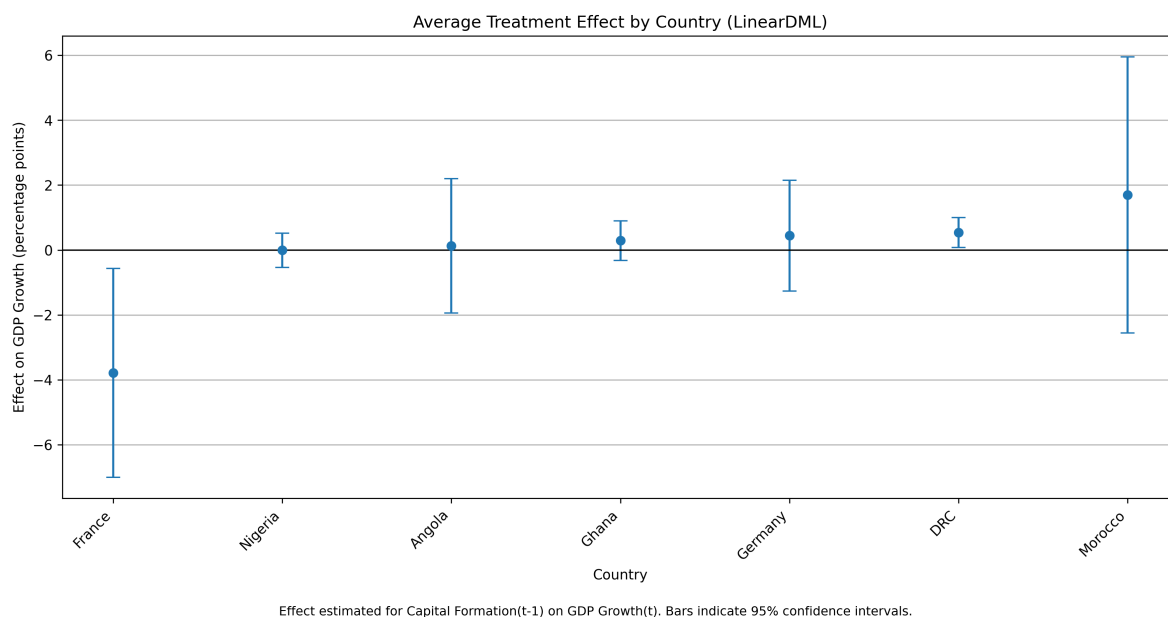


FIGURE 3.12 – Effets causaux moyens (ATE) par pays, LinearDML avec IC 95%.

Seuls la France ($p = 0.023$) et la RDC ($p = 0.022$) présentent des effets statistiquement significatifs, dans des directions opposées.

3.4.2.1 Interprétation Économique par Pays

France (ATE = -3.728 , $p = 0.023$) : L'effet négatif et significatif s'explique par deux mécanismes complémentaires. Premièrement, dans une économie à stock de capital mature, chaque unité d'investissement supplémentaire génère des rendements marginaux décroissants. Deuxièmement, en contexte de dette publique élevée (supérieure à 110% du PIB), l'investissement public peut exercer un effet d'éviction (*crowding-out*) sur l'investissement privé, en augmentant les taux d'intérêt réels [Mazumder et al. \(2025\)](#).

RDC (ATE = $+0.542$, $p = 0.022$) : L'effet positif et significatif reflète le déficit structurel de capital de cette économie, dont le taux d'investissement figure parmi les plus faibles du panel (cf. Figure 3.2). Dans ce contexte de rareté du capital physique, chaque unité d'investissement supplémentaire demeure productive. L'effet reste modéré ($+0.542$), ce qui peut refléter une capacité d'absorption institutionnelle limitée, facteur non contrôlé dans notre spécification (cf. Section 3.4.6) [\(Mulomba et al., 2025\)](#).

Nigeria, Angola, Ghana, Maroc (ATE non significatif) : L'absence d'effet significatif dans ces pays s'explique par la dominance d'autres canaux de croissance. Pour le Nigeria et l'Angola, les prix des matières premières (pétrole) déterminent principalement la croissance. L'investissement formel y reste peu déterminant. Pour le Maroc, les transferts de la diaspora et l'ouverture commerciale jouent un rôle prépondérant. Pour le Ghana, l'économie informelle (estimée à plus de 50% du PIB) implique que l'investissement officiel ne capte qu'une fraction des dynamiques réelles de croissance [Mulomba et al. \(2025\)](#); [Biaba et al. \(2025\)](#).

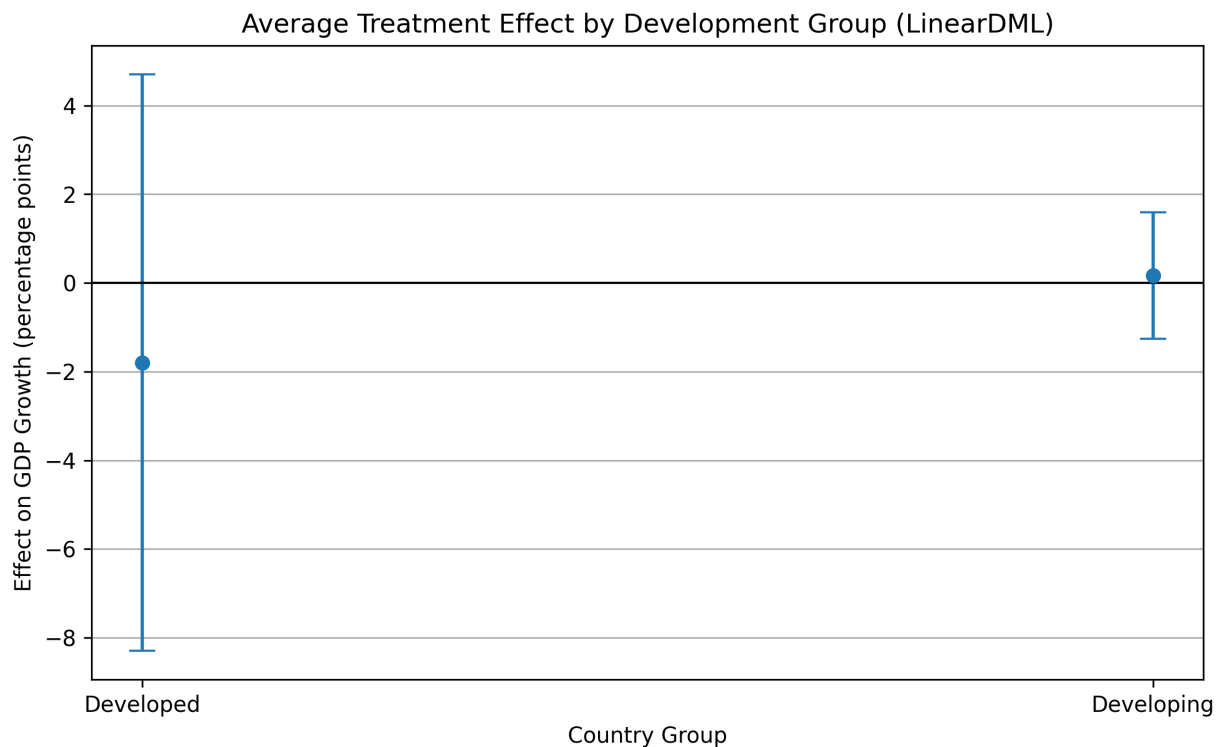
3.4.3 ATE par Groupe de Développement

Le Tableau 3.8 présente les résultats de l'estimation DML séparée pour les deux groupes de pays :

TABLE 3.8 – ATE par groupe (LinearDML avec BootstrapInference)

Groupe	n	ATE	IC 95%	p-val	Sig.
Développés	48	-1.797	[-8.294 ; +4.700]	0.588	Non
En développement	119	+0.167	[-1.261 ; +1.594]	0.819	Non

La Figure 3.13 présente les ATE estimés par LinearDML pour les deux groupes de développement, avec leurs intervalles de confiance à 95%.



Effect estimated for Capital Formation(t-1) on GDP Growth(t). Bars indicate 95% confidence intervals.

FIGURE 3.13 – ATE par groupe de développement avec LinearDML (IC 95%).

Les intervalles de confiance larges s'expliquent par les sous-échantillons limités (développés : $n = 48$; en développement : $n = 119$). La direction opposée des deux ATE confirme l'hétérogénéité structurelle entre groupes.

Les ATE par groupe ne sont pas individuellement significatifs en raison de ces sous-échantillons restreints (en particulier $n = 48$ pour les pays développés). La **direction opposée** des deux estimations est néanmoins cohérente avec les résultats par pays : effet négatif pour les pays développés (-1.797) et positif pour les pays en développement (+0.167).

3.4.4 Effets Hétérogènes Individuels : CausalForestDML

3.4.4.1 HTE par Groupe

Le Tableau 3.9 présente les effets de traitement hétérogènes estimés par CausalForestDML :

TABLE 3.9 – HTE par groupe avec CausalForestDML (800 arbres)

Groupe	n	ATE_CF	σ_{HTE}	Min / Max
Développés	48	-0.174	0.203	-0.577 / +0.200
En développement	119	+0.038	0.094	-0.224 / +0.264

Les deux estimateurs (LinearDML et CausalForestDML) convergent sur la direction de l'effet : négatif pour les pays développés, positif pour les pays en développement. La déviation standard plus élevée dans le groupe développé ($\sigma = 0.203$ contre 0.094) confirme une hétérogénéité interne plus forte, reflet des différences structurelles entre la France (effet très négatif) et l'Allemagne (effet légèrement positif).

3.4.4.2 Test de Différence Developed vs Developing

Le test de Welch appliqué aux distributions des effets individuels $\tau(x_i)$ produit les résultats suivants :

TABLE 3.10 – Test de Welch : Différence des HTE entre groupes

Paramètre	Valeur
Groupe A (Développés)	$n_A = 48$
Groupe B (En développement)	$n_B = 119$
Moyenne A	-0.174
Moyenne B	+0.038
t-statistique	-6.965
p-value	4.17×10^{-9}

La différence entre les deux groupes est hautement significative ($t = -6.965$, $p = 4.17 \times 10^{-9}$). Ce résultat constitue la contribution empirique centrale de ce travail : les effets de Capital_Formation sur GDP_Growth sont structurellement distincts entre pays développés et pays en développement, révélant deux régimes causaux opposés.

3.4.4.3 Distribution des Effets Individuels

La Figure 3.14 présente la distribution des effets de traitement hétérogènes individuels $\tau(x_i)$ estimés par CausalForestDML.

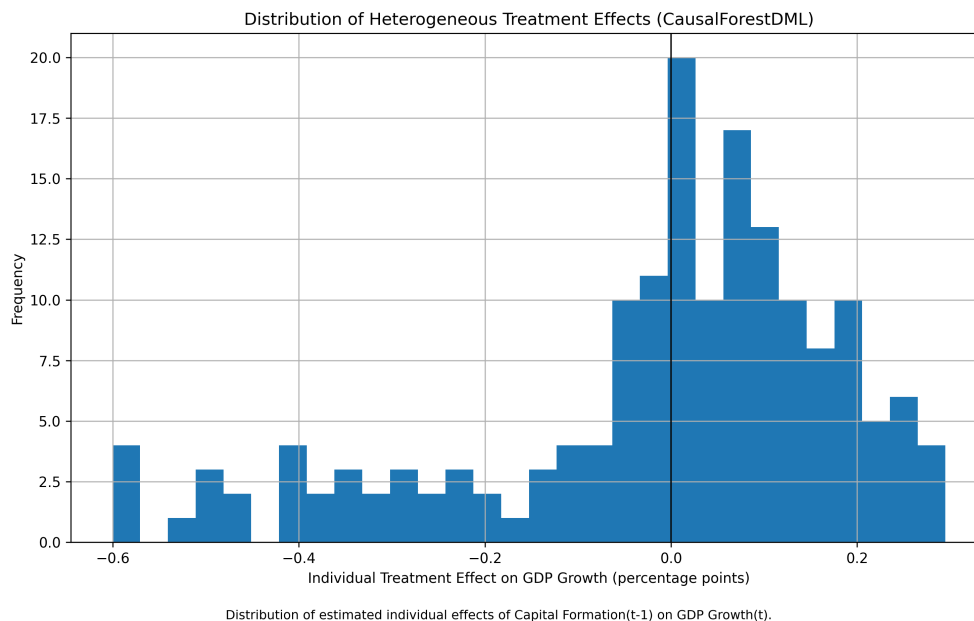


FIGURE 3.14 – Distribution des effets de traitement hétérogènes individuels $\tau(x_i)$ avec CausalForestDML.

La masse principale de la distribution est concentrée entre -0.1 et $+0.3$. La queue gauche (-0.6 à -0.2) correspond aux observations des pays développés (France, Allemagne). La distribution est asymétrique, avec un pic légèrement positif.

La distribution des effets individuels confirme l'hétérogénéité structurelle du panel. Les 10 effets les plus extrêmes en valeur absolue sont tous négatifs et appartiennent au groupe Développés : 8 observations pour l'Allemagne (entre -0.577 et -0.436) et 2 pour la France (entre -0.365 et -0.343). Ce résultat confirme que l'effet négatif global est **concentré sur les économies matures** du panel.

3.4.4.4 Comparaison LinearDML vs CausalForestDML

Le Tableau 3.11 compare les deux estimateurs par groupe :

TABLE 3.11 – Comparaison LinearDML vs CausalForestDML par groupe

Groupe	ATE LinearDML	ATE CausalForest	Convergence
Développés	-1.797	-0.174	Même signe (-)
En développement	$+0.167$	$+0.038$	Même signe (+)

La convergence des deux estimateurs sur le signe de l'effet renforce la robustesse des conclusions. La différence d'amplitude s'explique par la nature des deux approches. LinearDML estime un ATE moyen pondéré par le modèle linéaire, plus sensible aux valeurs extrêmes (notamment la France avec -3.728). CausalForestDML estime un effet médian local, moins influencé par les outliers [Oprescu et al. \(2019\)](#).

3.4.5 Discussion Générale des Résultats DML

Les résultats du Double Machine Learning répondent directement aux deux sous-problèmes posés en introduction.

Réponse au sous-problème 1 (structure causale) : Le Causal Discovery révèle que `Capital_Formation` n'est pas la cause directe et immédiate de la croissance du PIB dans les 7 pays du panel : son effet transite par des canaux intermédiaires, détaillés en Section 3.3.5.

Réponse au sous-problème 2 (quantification causale) : Le DML estime un ATE global de -0.528 ($p = 0.044$), négatif et significatif. Ce résultat moyen masque toutefois deux régimes causaux structurellement opposés. La France présente un effet fortement négatif (-3.728 , $p = 0.023$), cohérent avec les rendements décroissants du capital et l'effet d'éviction. La RDC affiche un effet positif significatif ($+0.542$, $p = 0.022$), cohérent avec son déficit structurel de capital. La différence entre les deux groupes est hautement significative ($p = 4.17 \times 10^{-9}$), confirmant l'existence de deux régimes causaux distincts.

Trois enseignements économiques majeurs ressortent de l'ensemble des résultats. L'effet de l'investissement sur la croissance n'est pas universel : il dépend du niveau de développement, du stock de capital existant et de la qualité institutionnelle du pays, et un modèle universel (*one-size-fits-all*) conduirait à des recommandations de politique économique erronées Biaba et al. (2025). Le Causal Machine Learning apporte une valeur ajoutée réelle par rapport à l'économétrie classique : la corrélation brute ($r = -0.02$) et le test de Granger (non significatif dans les 7 pays) suggèrent l'absence de relation, mais le DML révèle un effet causal significatif une fois les confondeurs correctement contrôlés.

3.4.6 Discussion de la robustesse des estimations

La robustesse des estimations DML repose sur plusieurs éléments vérifiables à partir des résultats obtenus. Premièrement, la convergence entre les deux estimateurs retenus est le LinearDML et CausalForestDML sur la direction des effets (négatif pour les pays développés, positif pour les pays en développement) constitue un indicateur de cohérence interne. Une divergence de signe entre les deux estimateurs aurait signalé une instabilité structurelle de l'identification causale.

Deuxièmement, la procédure de cross-fitting à 5 plis, répétée avec différentes initialisations aléatoires, produit des estimations stables en termes de signe et d'ordre de grandeur pour les pays présentant des effets significatifs (France et RDC). Cette stabilité sur plusieurs répliques constitue une forme de robustesse computationnelle, bien qu'elle ne remplace pas des tests de permutation formels.

Troisièmement, la cohérence entre les résultats du Causal Discovery et ceux du DML renforce la validité des conclusions. L'absence de lien direct Formation de Capital $\rightarrow$ PIB identifiée par les quatre méthodes de découverte causale est cohérente avec l'ATE global quasi nul (-0.528) masquant deux régimes opposés, plutôt qu'avec un effet positif et universel que la théorie standard prédirait naïvement.

Une limite d'identification mérite d'être explicitée. L'absence de contrôle institutionnel dans la matrice des confondeurs implique que les effets estimés peuvent absorber une part de la variation attribuable à la qualité de la gouvernance. Le sens probable de ce biais diffère selon les groupes : dans les économies à gouvernance fragile, une amélioration institutionnelle tend à accroître simultanément l'investissement et la croissance, ce qui conduirait à surestimer l'effet causal du capital. À l'inverse, dans les économies matures où la qualité institutionnelle varie peu, ce biais est vraisemblablement faible. L'effet positif estimé pour la RDC doit donc être lu comme une borne haute plutôt que comme une estimation ponctuelle, tandis que l'effet négatif estimé pour la France apparaît moins exposé à

cette source de biais.

Plusieurs tests de robustesse complémentaires restent à conduire dans une étape ultérieure : un test placebo consistant à permuer aléatoirement les valeurs du traitement et à vérifier que l'ATE estimé converge vers zéro [Wager and Athey \(2018\)](#) ; une analyse de sensibilité omettant successivement chaque confondeur pour évaluer leur contribution individuelle au contrôle du biais ; et une validation externe sur un sous-panel de pays non utilisés dans l'estimation. Ces éléments feront l'objet du second article en cours de soumission.

3.5 Conclusion partielle

Ce chapitre a présenté et analysé les résultats expérimentaux obtenus en deux phases complémentaires.

La Phase 1 de Causal Discovery a révélé que le lien `Capital_Formation` $\rightarrow$ `GDP_Growth` est absent ou fragile dans les 7 pays du panel, selon les 4 méthodes utilisées. Le taux de change s'impose comme la variable causale dominante, et la dette publique comme le confondeur le plus robuste. Ces résultats confirment que l'effet de l'investissement sur la croissance est **indirect et médiatisé** par d'autres variables macroéconomiques, notamment dans les économies africaines où la contrainte extérieure est déterminante [Mulomba et al. \(2025\)](#).

La Phase 2 de Double Machine Learning a quantifié cet effet causal une fois les confondeurs rigoureusement contrôlés. L'ATE global est négatif et significatif (-0.528 , $p = 0.044$). Ce résultat moyen masque deux régimes causaux opposés. Un effet négatif significatif s'observe en France (-3.728 , $p = 0.023$), cohérent avec les rendements décroissants du capital dans les économies matures. Un effet positif significatif s'observe en RDC ($+0.542$, $p = 0.022$), cohérent avec le déficit structurel de capital. La différence entre pays développés et pays en développement est hautement significative ($p = 4.17 \times 10^{-9}$), constituant le résultat empirique central de ce travail.

Quatre contributions majeures ressortent de ce chapitre. La complémentarité entre Causal Discovery et DML est démontrée : le premier identifie les structures causales, le second les quantifie. L'hétérogénéité structurelle entre régimes de développement est mise en évidence, confirmant l'absence d'effet universel de l'investissement sur la croissance. La faisabilité technique du CML sur données macroéconomiques africaines est démontrée, malgré les contraintes de séries courtes et d'hétérogénéité structurelle.

Le chapitre suivant présente les conclusions générales de ce travail, ses limites et les perspectives de recherche futures.

Conclusion Générale et Perspectives

Synthèse des Contributions

À notre connaissance, ce travail constitue l’une des premières applications empiriques combinant Causal Discovery et Double Machine Learning sur un panel macroéconomique hétérogène de 7 pays (France, Allemagne, Maroc, Ghana, Nigeria, Angola, RDC) sur 25 années (2000–2024). Sur le plan méthodologique, le pipeline proposé est reproductible et directement transposable à d’autres contextes de panel macroéconomique. Sur le plan empirique, deux effets causaux significatifs ont été identifiés, un effet négatif en France ($ATE = -3.728$, $p = 0.023$) et un effet positif en RDC ($ATE = +0.542$, $p = 0.022$) — révélant deux régimes causaux structurellement distincts confirmés par le test de Welch ($p = 4.17 \times 10^{-9}$). Ces résultats démontrent que l’analyse causale par CML est techniquement faisable sur des données macroéconomiques africaines à séries courtes, avec des adaptations méthodologiques appropriées.

Réponse à la Problématique

La problématique centrale posée en introduction était :

Est-il possible, à l’aide des méthodes de Causal Machine Learning, d’identifier et de quantifier l’effet causal net de la formation de capital (en $t - 1$) sur la croissance du PIB (en t), en contrôlant rigoureusement les confondeurs macroéconomiques, et cet effet varie-t-il de manière structurelle selon le niveau de développement économique du pays ?

La réponse est **oui** sur les deux volets. Il est possible d’identifier et de quantifier cet effet causal avec le CML. Cet effet varie par ailleurs de manière structurelle et hautement significative selon le niveau de développement : négatif dans les économies matures (rendements décroissants, effet d’éviction) et positif dans les économies à déficit de capital (rendements croissants, absorption productive de l’investissement).

Limites du Travail

Malgré les contributions apportées, ce travail comporte plusieurs limites. La taille des sous-échantillons constitue la première limite : avec seulement 24 observations par pays, la puissance statistique des estimations par pays est limitée, les intervalles de confiance sont larges pour plusieurs pays (Angola, Maroc, Allemagne), et seuls 2 pays sur 7 présentent des effets significatifs au seuil de 5%. L’hypothèse d’unconfoundedness pose une seconde limite : des variables institutionnelles cruciales comme la qualité de gouvernance, l’indice de corruption ou le niveau d’État de droit sont absentes du panel et constituent des confondeurs potentiellement non observés. Le secteur informel

représente une troisième limite, particulièrement critique pour la RDC et le Ghana où il représente plus de 50% du PIB et échappe aux indicateurs officiels du FMI et de la Banque Mondiale. L'hypothèse SUTVA de non-interférence entre unités est contestable dans un contexte de globalisation où les économies sont interdépendantes. Enfin, les chocs exogènes majeurs (prix du pétrole, conflits armés, pandémie COVID-19) sont traités comme bruit résiduel dans le modèle, alors qu'ils peuvent constituer des déterminants majeurs de la croissance, notamment pour le Nigeria et l'Angola.

Perspectives

Les résultats et les limites identifiées ouvrent plusieurs directions de recherche prometteuses. L'extension à des données trimestrielles augmenterait le nombre d'observations de $T = 25$ à $T \approx 100$ par pays, renforçant la puissance statistique des estimations et permettant une modélisation plus fine des dynamiques de transmission. L'inclusion d'indicateurs de gouvernance (Worldwide Governance Indicators, Indice de Perception de la Corruption) comme confondeurs supplémentaires permettrait de tester si l'effet de l'investissement est conditionnel à la qualité institutionnelle, hypothèse centrale dans la littérature sur la croissance en Afrique. L'application du Double Machine Learning dynamique (DDML) permettrait de capturer les effets inter-temporels de l'investissement sur plusieurs périodes, au-delà du lag d'un an retenu ici. L'extension du panel à 15–20 pays africains supplémentaires permettrait de généraliser les résultats et d'identifier des sous-groupes régionaux homogènes aux régimes causaux distincts. L'utilisation d'une variable instrumentale (investissements directs étrangers, aide publique au développement) renforcerait l'identification causale au-delà de l'hypothèse d'unconfoundedness. Enfin, les résultats de ce stage constituent la base d'un second article scientifique en cours de rédaction, venant compléter la revue systématique publiée à EUSPN 2025.

Conclusion

Ce travail de stage démontre que le Causal Machine Learning constitue une avancée méthodologique réelle pour l'analyse macroéconomique, particulièrement dans des contextes hétérogènes combinant économies développées et économies africaines. En révélant des régimes causaux structuellement distincts que l'économétrie classique ne pouvait pas identifier, le CML ouvre de nouvelles perspectives pour la conception de politiques économiques différenciées et contextuellement adaptées.

Le message central de ce travail est clair : il n'existe pas d'effet universel de l'investissement sur la croissance. La politique économique doit tenir compte du niveau de développement, du stock de capital existant et de la qualité institutionnelle de chaque économie pour être efficace. Le Causal Machine Learning fournit les outils nécessaires pour identifier et quantifier cette hétérogénéité avec rigueur. Il ouvre ainsi la voie à une politique économique fondée sur l'évidence causale, et non sur la simple corrélation.

Bibliographie

- Biaba, J. K., Mulomba, C. M., Bazie, I. G., Nzazi, B. N., Tashev, T., Kambale, W. V., Kyamakya, K., Kasoro, N. M., Tuong, V. H., and Kasereka, S. K. (2025). Causal machine learning applied to macroeconomic analysis : Applications, challenges, and perspectives. *Procedia Computer Science*, 272 :251–260.
- Bond, S. R., Hoeffler, A., and Temple, J. R. W. (2001). GMM estimation of empirical growth models. CEPR Discussion Paper 3048, Centre for Economic Policy Research, London. Également paru comme Economics Papers 2001-W21, Nuffield College, University of Oxford.
- Calderón, C. and Servén, L. (2008). Infrastructure and economic development in sub-Saharan Africa. Policy Research Working Paper 4712, The World Bank, Washington, DC.
- Chen, X., Wohlfarth, P., and Smith, R. P. (2021). China’s money demand in a cointegrating vector error correction model. *Journal of Asian Economics*, 75 :101338.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.
- Cordoni, F. and Sancetta, A. (2024). Consistent causal inference for high-dimensional time series. *Journal of Econometrics*, 246(1–2) :105902.
- Guo, R., Cheng, L., Li, J., Hahn, P. R., and Liu, H. (2020). A survey of learning causality with data : Problems and methods. *ACM Computing Surveys*, 53(4) :1–37.
- Krich, C., Runge, J., Miralles, D. G., Migliavacca, M., Perez-Priego, O., El-Madany, T., Carrara, A., and Mahecha, M. D. (2020). Estimating causal networks in biosphere-atmosphere interaction with the PCMCI approach. *Biogeosciences*, 17(4) :1033–1061.
- Liu, Y., Wang, Y., Zheng, G., Wang, J., and Guo, K. (2019). The dynamical relationship between capital market and macroeconomy : Based on dynamic Bayesian network. *Procedia Computer Science*, 162 :46–52.
- Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1) :3–42.
- Maitra, S. (2025). Vector auto regression and vector error correction model. In *A Practical Guide to Static and Dynamic Econometric Modelling : Examples and Analysis with Python Code Embedded*, pages 171–201. Springer.

- Mazumder, M. T. R., Shourov, M. S. H., Rasul, I., Akter, S., and Miah, M. K. (2025). The impact of macroeconomic factors on the US market : A data science perspective. *Journal of Economics, Finance and Accounting Studies*, 7(2) :208–219.
- Mulomba, C. M., Kiketa, V. M., Kutangila, D. M., Mampuya, P. H. K., Mukenze, N., Kasunzi, L. M., Kyamakya, K., Tashev, T., and Kasereka, S. K. (2025). Applying causal machine learning to spatio-temporal data analysis : An investigation of opportunities and challenges. *IEEE Access*.
- Olayungbo, D. O. (2019). Bayesian graphical model application for monetary policy and macroeconomic performance in Nigeria. In *Bayesian Networks : Advances and Novel Applications*. IntechOpen.
- Oprescu, M., Syrgkanis, V., Battocchi, K., Hei, M., and Lewis, G. (2019). EconML : A machine learning library for estimating heterogeneous treatment effects. In *Proceedings of the 33rd Conference on Neural Information Processing Systems*.
- Pearl, J. (2009). *Causality : Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2 edition.
- Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica*, 56(4) :931–954.
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5) :S71–S102.
- Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects. *Statistics in Medicine*, 26(1) :20–36.
- Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. (2019). Inferring causation from time series in Earth system sciences. *Nature Communications*, 10 :2553.
- Shojaie, A. and Fox, E. B. (2022). Granger causality : A review and recent advances. *Annual Review of Statistics and Its Application*, 9 :195–220.
- Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1) :65–94.
- Strobl, E. V., Visweswaran, S., and Spirtes, P. L. (2018). Fast causal inference with non-random missingness by test-wise deletion. *International Journal of Data Science and Analytics*, 6(1) :47–62.
- Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523) :1228–1242.
- Zhao, Y. and Liu, Q. (2023). CausalML : Python package for causal inference machine learning. *SoftwareX*, 21 :101294.

QUYẾT ĐỊNH

Về việc thành lập Hội đồng đánh giá luận văn thạc sĩ

GIÁM ĐỐC ĐẠI HỌC QUỐC GIA HÀ NỘI

Căn cứ Nghị định số 201/2025/NĐ-CP ngày 11 tháng 07 năm 2025 của Chính phủ quy định chức năng, nhiệm vụ và quyền hạn của đại học quốc gia;

Căn cứ Quyết định số 3636/QĐ-ĐHQGHN của Giám đốc Đại học Quốc gia Hà Nội ngày 21 tháng 10 năm 2022 của Giám đốc Đại học Quốc gia Hà Nội ban hành Quy chế đào tạo thạc sĩ tại Đại học Quốc gia Hà Nội;

Theo đề nghị của Hiệu trưởng Trường Quốc tế tại công văn số 1110/TQT-ĐT ngày 24 tháng 06 năm 2026 về việc đề xuất thành lập hội đồng đánh giá luận văn cho học viên;

Theo đề nghị của Trưởng ban Đào tạo và Công tác sinh viên.

QUYẾT ĐỊNH:

Điều 1. Thành lập Hội đồng đánh giá luận văn thạc sĩ của học viên cao học BIABA KUYA Jirince, sinh ngày 27/01/1994, tại Công-gô;

Đề tài: Phân tích nhân quả về tác động kinh tế vĩ mô từ chuỗi thời gian bằng Học máy nhân quả;

Chuyên ngành: Hệ thống thông minh và Đa phương tiện; Mã số: 8480201.02;

Danh sách các thành viên của Hội đồng kèm theo Quyết định này.

Điều 2. Hội đồng có trách nhiệm đánh giá luận văn thạc sĩ theo Quy chế đào tạo thạc sĩ tại Đại học Quốc gia Hà Nội và tự giải thể sau khi hoàn thành nhiệm vụ.

Điều 3. Chánh Văn phòng, Trưởng ban Đào tạo và Công tác sinh viên, Hiệu trưởng Trường Quốc tế và các thành viên trong Hội đồng đánh giá luận văn thạc sĩ chịu trách nhiệm thi hành Quyết định này.

Nơi nhận:

- Như Điều 3;
- Giám đốc (để báo cáo);
- Lưu: VT, ĐT&CTSV, M10.

KT. GIÁM ĐỐC
PHÓ GIÁM ĐỐC


Đào Thanh Trường

DANH SÁCH HỘI ĐỒNG ĐÁNH GIÁ LUẬN VĂN THẠC SĨ

*(Kèm theo Quyết định số: 3219 /QĐ-ĐHQGHN, ngày 08 tháng 7 năm 2026
của Giám đốc Đại học Quốc gia Hà Nội)*

STT	Họ và tên	Ngành/ Chuyên ngành	Cơ quan công tác	Trách nhiệm trong Hội đồng
1	TS. Nguyễn Thị Minh Huyền	Công nghệ thông tin	Trường Đại học Khoa học Tự nhiên, ĐHQGHN	Chủ tịch
2	TS. Sonfack Souchio Serge	Công nghệ thông tin	Đại học Joseph Ki- ZERBO, Burkina Faso	Phản biện 1
3	TS. Hồ Tường Vinh	Công nghệ thông tin	Trường Quốc tế, ĐHQGHN	Phản biện 2
4	TS. Vũ Thị Hương Giang	Công nghệ thông tin	Đại học Bách khoa Hà Nội	Thư ký
5	TS. Nguyễn Hồng Quang	Công nghệ thông tin	Hội Tin học Việt Nam (VAIP)	Ủy viên

Hội đồng gồm 05 thành viên.

Hanoi, le 21/07/2026

RESOLUTION DU JURY D'ÉVALUATION DE MÉMOIRE DE MASTER

Vu la Décision n°3219/QĐ-ĐHQGHN émise le 08/07/2026 par le Président de l'Université nationale du Vietnam, Hanoï, portant création du jury d'évaluation de mémoire de master pour évaluer le mémoire de master de l'étudiante **BIABA KUYA Jirince** (date de naissance : 27/01/1994)

Titre du mémoire : Analyse causale de l'impact macroéconomique à partir de séries temporelles à l'aide du Causal Machine Learning g

Spécialité : Systèmes intelligents et multimédia

Code : 8480201.02

Le jury d'évaluation de mémoire s'est réuni le 21/07/2026 à l'École internationale de l'Université nationale du Vietnam, Hanoï.

Après la présentation du mémoire par l'étudiant(e), des observations des rapporteurs, les commentaires des membres du jury et les réponses de l'étudiant(e), le jury a délibéré et a rendu sa décision:

1. La nécessité, la signification scientifique et l'aspect pratique du sujet de mémoire

Le sujet est actuel et pertinent. Il applique le Causal Machine Learning à l'analyse de l'effet de l'investissement sur la croissance et présente un intérêt particulier pour les économies africaines. Les résultats peuvent fournir des indications utiles aux décideurs, mais ne permettent pas encore une application directe aux politiques publiques.

2. Structure, méthodologie de recherche, références, etc., du mémoire

Le mémoire est bien structuré et s'appuie sur une bibliographie adaptée. La méthodologie combine Granger, VAR, VECM, PCMCI, LinearDML et CausalForestDML. Certains choix, notamment la construction du DAG de consensus et le traitement des séries courtes, doivent cependant être mieux justifiés.

3. Résultats de la recherche:

Le travail analyse sept pays sur la période 2000–2024. Il montre que le lien direct entre formation de capital et croissance reste faible, tandis que le taux de change joue un rôle

important. Le DML révèle notamment des effets significatifs et opposés en France et en RDC.

4. Limitations et demandes de correction du mémoire (le cas échéant)

L'échantillon reste faible, avec environ 24 observations par pays. Il convient de vérifier le score de support du Nigeria, de mieux analyser les biais des données, de comparer les résultats avec des modèles économétriques traditionnels et de nuancer les conclusions causales.

5. Évaluation globale:

Le candidat a réalisé un travail sérieux et ambitieux sur un sujet actuel, en mobilisant des méthodes avancées de Causal Machine Learning. Le travail présente un intérêt méthodologique et a déjà donné lieu à des productions scientifiques.

Toutefois, les résultats restent exploratoires en raison de la taille limitée des données. Certains éléments méthodologiques, notamment la validation économique, la comparaison avec les modèles traditionnels et le calcul des scores de support, doivent être clarifiés ou corrigés. Le mémoire est globalement jugé satisfaisant, sous réserve de la prise en compte des remarques du jury.

Note du mémoire: 8.6/10.

Résultat de la soutenance de mémoire de master de l'étudiant BIABA KUYA Jirince:

Admis : x

Non admis : ☐

(Le mémoire est admis si la note moyenne attribuée par le jury est de 5,5 ou plus).

Cette résolution est adoptée à l'unanimité par 5/5 membres du jury.

SECRETAIRE DU JURY

PRESIDENT DU JURY



Dr. Vũ Thị Hương Giang



Dr. Nguyễn Thị Minh Huyền

**Certification de l'Ecole innationale
P.P. RECTEUR
DEPARTEMENT DE SCOLARITE**



Prof. Assoc. Dr. Hoàng Tuyết Minh

BẢN XÁC NHẬN/ATTESTATION
Về việc chỉnh sửa luận văn thạc sĩ sau bảo vệ
De la correction du mémoire après la soutenance

Họ và tên/*Nom et prénom*: **BIABA KUYA Jirince**
Lớp/*Classe*: P27 SIM
Khóa/*Promotion*: QH - 2023
Chuyên ngành/*Parcours*: Hệ thống thông minh và Đa phương tiện/*Systèmes Intelligents et Multimédia*
Ngày bảo vệ/*Date de soutenance*: 21/07/2026
Đề tài luận văn/*Sujet de mémoire*:
- Tiếng Việt/*En vietnamien*: Phân tích nhân quả về tác động kinh tế vĩ mô từ chuỗi thời gian bằng Học máy nhân quả
- Tiếng Pháp/*En français*: *Analyse causale de l'impact macroéconomique à partir de séries temporelles à l'aide du Causal Machine Learning*
Giáo viên hướng dẫn/*Encadrant*: Professeur Selain KASERKA KABUNGA
Thời gian bảo vệ luận văn/*Date de soutenance*: 21/07/2026
Sau khi bảo vệ luận văn, được sự góp ý của Hội đồng chấm luận văn thạc sĩ, tôi đã chỉnh sửa và hoàn chỉnh luận văn theo các vấn đề sau/*Après avoir soutenu le mémoire, avec des remarques du Jury d'évaluation du mémoire, j'ai corrigé et achevé mon mémoire pour les points ci-dessous*:

STT	Nội dung góp ý của hội đồng <i>Remarques du Jury</i>	Nội dung đã chỉnh sửa <i>Contenu corrigé</i>	Mục/Trang đã chỉnh sửa/ <i>Chapitre/page corrigé(e)</i>
1	Sections 1.4.2, 1.4.3 et 1.7.1 : ajouter la description des variables GDP_Growth et Capital_Formation	Description explicite des deux variables ajoutée aux trois emplacements, avec renvoi au Tableau 2.1	§1.4.2, §1.4.3, §1.7.1 : p. 11, p.13
2	Section 1.12.1 : références manquantes et formatage	Ajout de la référence Lucas (1988) ; passage des citations en \citet\citep ; correction du formatage dans tout le document	§1.12.1: p. 18
3	Page 28 : Figure 2.2 non disponible / doublon avec Figure 3.14	Figure 3.14 supprimée (contenu redondant) ; affichage de la Figure 2.2 vérifié	§3.4.4.3: p. 48

4	Page 31 : les valeurs de la Figure 3.1 ne correspondent pas à la description	Légendes et commentaires des Figures 3.1 et 3.2 harmonisés avec l'échelle standardisée (Z-score)	§3.2.1 : p. 31
5	Page 33 : la corrélation de la Figure 3.3 ne correspond pas au texte	Valeur corrigée en $r = -0.02$ dans les six occurrences du document	§1.7.1, §2.3.4, §2.4.1, §3.2.2, §3.2.3, §3.4.5
6	Manque de variables de contrôle institutionnelles	Limite explicitée en §2.6.2 ; analyse du sens du biais ajoutée en §3.4.6 ; interprétation du résultat RDC ajustée	§2.6.2, §3.4.2.1, §3.4.6 : p. 27, p.46, p.50
7	Répétitions théoriques du Chapitre 3 et justification méthodologique du Chapitre 2	Suppression des passages redondants en §3.3.4.2, §3.4.5, §3.5 et §1.13.3	Ch. 1 et 3 : p. 40, p.49, p.51, p.19

Hà Nội, ngày 23 tháng 07 năm 2026

XÁC NHẬN CỦA GIÁO VIÊN HƯỚNG DẪN
APPROBATION DE L'ENCADRANT

Prof. Selain KASEREKA KABUNGA

HỌC VIÊN
ÉTUDIANT

BIABA KUYA Jirince

XÁC NHẬN CỦA CHỦ TỊCH HỘI ĐỒNG
APPROBATION DU PRÉSIDENT DU JURY

Nguyễn Thị Minh Huyền

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc/
SOCIALIST REPUBLIC OF VIETNAM
Independence – Freedom – Happiness

GIẤY THỎA THUẬN/ AGREEMENT LETTER
V/v Sử dụng, khai thác nội dung tài liệu nộp lưu chiểu/
Regarding the use and exploitation of submitted archival documents

Tôi tên là/ *Full name*: BIABA KUYA JIRINCE Số điện thoại/ *Tel*: +243820031916

Là tác giả tài liệu/*As the author of the document*:

Causal Analysis of Macroeconomic Impact from Time Series using Causal Machine Learning

Tôi làm bản thỏa thuận này với Thư viện Trường Quốc tế như sau/ *I make this agreement with the VNU-IS Library as follows*:

Tôi đồng ý cho phép Thư viện Trường Quốc tế sử dụng và khai thác nội dung tài liệu vào mục đích phục vụ nghiên cứu khoa học, giảng dạy và học tập trong ĐHQGHN/ *I agree to allow the VNU-IS Library to use and exploit the content of the documents for the purposes of scientific research, teaching, and learning within the Vietnam National University./.*

Hà Nội, ngày 01 tháng 08 năm 2026

Tác giả/ *Author*

BIABA KUYA JIRINCE



(Ký và ghi rõ họ tên/*Full name and signature*)