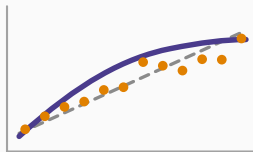


Économétrie avancée

Chapitre 19 : Réaliser un projet empirique



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Avant de commencer

Poser une question précise

Revue de littérature et collecte de données

Analyse économétrique

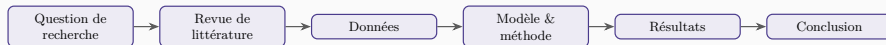
Rédiger le rapport

Synthèse

Avant de commencer

Pourquoi ce chapitre ?

Dix-sept chapitres ont accumulé des outils économétriques. Ce chapitre change de registre : il s'agit de méthodologie de recherche appliquée — comment transformer une question vague en projet empirique mené à terme, de la formulation de la question jusqu'à la rédaction du rapport final. Ce n'est pas de la théorie supplémentaire, mais un guide pratique directement utilisable pour un mémoire, un projet de fin d'études ou une note de recherche.



Poser une question précise

De l'idée vague à la question testable

« Je travaille sur la criminalité » n'est pas une question de recherche ; « j'étudie l'effet de la police de proximité sur le taux de criminalité urbaine aux États-Unis » en est une. Une question précise détermine directement quelles variables collecter, sur quelle population, et sur quelle période — s'en dispenser conduit typiquement à découvrir, trop tard, qu'une variable clé n'a pas été recueillie ou que l'échantillon ne correspond pas à la population visée.

Ambition et faisabilité

Un projet d'un semestre ne peut viser l'originalité totale sur des agrégats macroéconomiques déjà abondamment étudiés (courbe de Phillips américaine, fonction de consommation agrégée). Les gains marginaux les plus probables viennent de trois sources : un pays ou une période moins étudiés, une nouvelle variable explicative, ou un jeu de données mieux construit (panel plutôt que coupe transversale, meilleures variables proxy) pour une question déjà classique.

Revue de littérature et collecte de données

Revue de littérature

Même un travail court doit situer son apport par rapport aux précédents. Les bases de données bibliographiques spécialisées et les moteurs de recherche par mots-clés, titre ou auteur permettent d'identifier rapidement les articles pertinents; il est utile d'élargir la recherche à des thèmes connexes non couverts par les mots-clés évidents (ex. : étudier l'usage de drogues sur la performance scolaire suggère de consulter aussi la littérature sur l'alcool).

Choisir la structure de données adaptée

Coupe transversale, série temporelle, coupes répétées ou panel : le choix dépend de la question et de ce qui est disponible. Un panel permet de contrôler les effets inobservés constants dans le temps (chapitres 13-14) — un avantage décisif quand l'analyse en coupe transversale simple laisse planer un doute sérieux sur l'endogénéité (auto-sélection, simultanéité).

Organiser et nettoyer les données

Règles pratiques essentielles : chaque observation en ligne, chaque variable en colonne ; pour les séries temporelles, ordre chronologique strict avec un indicateur de période ; pour un panel, années consécutives regroupées par unité avec un identifiant unique. Les codes de valeurs manquantes numériques (ex. -99, 999) sont une source classique d'erreurs silencieuses si le logiciel les traite comme des observations valides — toujours vérifier les minimums, maximums, moyennes de chaque variable avant l'estimation.

Analyse économétrique

Justifier les hypothèses d'identification

Utiliser les MCO exige d'argumenter que le terme d'erreur n'est pas corrélé aux régresseurs retenus : l'auto-sélection (chapitres 7 et 17), l'erreur de mesure et la simultanéité (chapitres 9 et 16) sont les trois sources d'endogénéité à écarter explicitement, pas seulement à mentionner. Si une variable instrumentale est utilisée, ses deux conditions — exclusion de l'équation structurelle et corrélation partielle testable avec la variable endogène — doivent être discutées séparément (chapitre 15).

L'analyse de sensibilité, signature d'un travail sérieux

Un bon travail empirique fait varier raisonnablement la spécification — remplacer une mesure continue par une indicatrice, exclure les valeurs aberrantes, changer la forme fonctionnelle — et rapporte si les conclusions centrales résistent. À l'inverse, rechercher systématiquement la spécification qui produit un résultat significatif (régression pas à pas automatisée ou recherche manuelle répétée) invalide l'interprétation des tests t et F obtenus : c'est le **dragage de données** (*data mining*), à éviter ou du moins à ne jamais dissimuler.

Rédiger le rapport

Les sections d'un travail empirique

Introduction (question, enjeu, aperçu des résultats); cadre conceptuel (théorie ou raisonnement économique justifiant les variables retenues); modèle économétrique et méthode d'estimation (équation de population, sans « chapeaux » sur les coefficients — un modèle n'est pas une méthode d'estimation); données (sources, définitions, statistiques descriptives); résultats (interprétation, significativité économique **et** statistique); conclusion (limites, pistes futures).

Modèle versus méthode d'estimation

Une confusion fréquente : « un modèle MCO » n'a pas de sens. Le **modèle** est une relation de population, par exemple $\text{colGPA} = \beta_0 + \beta_1 \text{alcohol} + \beta_2 \text{hsGPA} + u$; les **MCO**, le **2SLS** ou les **MCG faisables** sont des méthodes permettant de l'estimer. La section « modèle » présente l'équation sans anticiper les résultats (ne pas justifier l'omission d'une variable par son insignifiance constatée *a posteriori* — cela appartient à la section résultats).

Bonnes pratiques de présentation

Toujours indiquer le nombre d'observations et le R^2 pour chaque régression rapportée; choisir les unités de mesure pour éviter des coefficients illisibles (notation scientifique à proscrire); limiter le nombre de décimales affichées; ne jamais inclure de commandes logicielles dans le corps du texte. Un tableau de statistiques descriptives (moyenne, écart-type, minimum, maximum) et un tableau de définition des variables, cohérents en nomenclature avec les équations présentées, facilitent grandement la lecture.

Synthèse

L'essentiel du chapitre 18

- Une question de recherche précise, formulable en une phrase, conditionne tout le reste du projet.
- La structure des données (coupe, série temporelle, panel) doit être choisie en fonction de la capacité à isoler un effet *ceteris paribus* crédible.
- Toute hypothèse d'identification (exogénéité, validité d'un instrument) doit être argumentée explicitement, pas simplement supposée.
- L'analyse de sensibilité renforce la crédibilité d'un résultat; la recherche répétée de spécifications significatives (dragage de données) la détruit.
- Un rapport empirique suit une structure standard : question, cadre conceptuel, modèle et méthode, données, résultats, conclusion — avec une présentation soignée des tableaux et des unités.

Vers le chapitre 19

Le dernier chapitre du programme revient aux fondations : un rappel structuré des outils mathématiques de base (fonctions, dérivées, optimisation élémentaire) qui sous-tendent l'ensemble de l'appareil économétrique présenté depuis le chapitre 1.