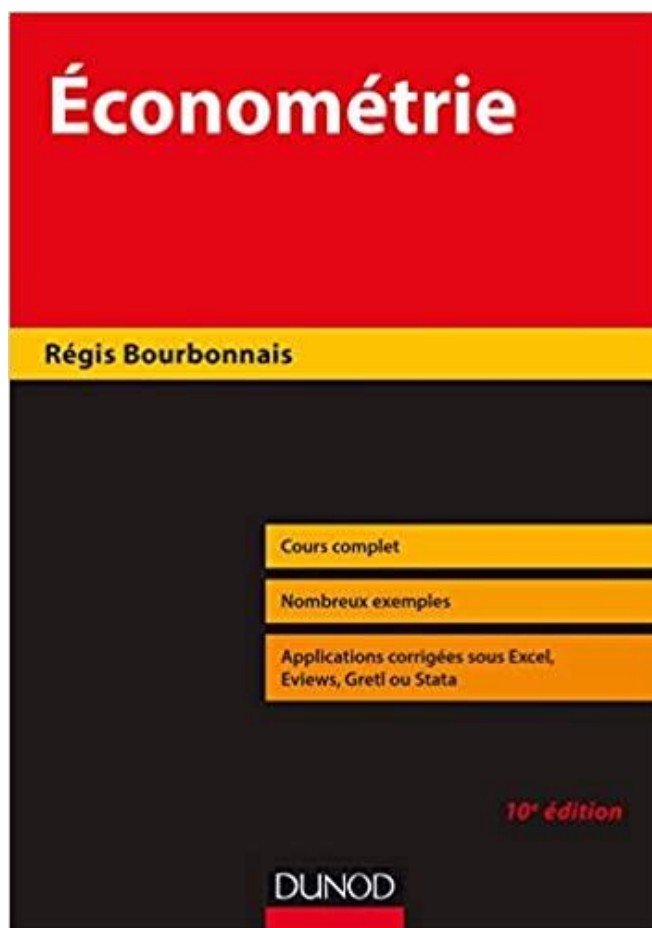




## Econométrie (10<sup>e</sup> édition)

Régis Bourbonnais

Dunod, 2018.



Un beau jour, il y a bien des années de cela, le responsable de notre L3 (licence 3<sup>ème</sup> année) déboule dans mon bureau et m'annonce tout de go : « un tel lâche son cours d'économétrie, on est mal, je sais que tu as fait des études d'économétrie, est-ce que ça te dit de le prendre ? ». J'étais pris de court, c'est le moins qu'on puisse dire. Certes, j'avais suivi des études d'économétrie – jusqu'en DEA, l'équivalent d'un Master Recherche – mais, comme tout bon étudiant, surtout de mon époque, j'avais avant tout appris pour reproduire en vue de l'examen.

Enseigner est une tout autre histoire. Je m'en étais rendu compte. A ce moment-là, je pratiquais depuis plusieurs années déjà. Au fil du temps, je m'étais spécialisé dans les cours de programmation, de machine learning et ses applications. Je prenais de plus en plus de recul par rapport à ce que je transmettais. Mon point de vue sur l'économétrie en revanche restait très scolaire. Je ne pouvais décemment pas me présenter devant les étudiants, dicter un cours insipide en ânonnant des choses dites et redites partout dans les ouvrages, ou que l'on retrouve



facilement en effectuant quelques requêtes sous Google, bien qu'au début / milieu des années 2000, les ressources n'étaient pas aussi abondantes qu'aujourd'hui.

Après quelques hésitations, j'ai accepté. C'est le genre de challenge qui me stimule. Il m'a donc fallu [préparer le module](#). Le contenu était relativement bien cadré. En licence, il s'agissait de revenir rapidement sur l'analyse de corrélation et la régression simple, puis explorer la régression multiple (je me cantonnais aux données transversales), en traitant les aspects pratiques, entre autres la sélection de variables, l'étude des points atypiques, les exogènes qualitatives. Dans un réflexe atavique, j'ai ressorti mes anciennes notes de cours. J'ai très rapidement compris que ce n'était pas la bonne idée. Il faut que je puisse mener les séances en les adaptant à mon style (j'avoue que j'aime bien quand ça pulse) et aux caractéristiques des étudiants d'aujourd'hui (ils ont besoin que ça pulse aussi sinon ils s'ennuient, ça tombe bien pour le coup).

J'en ai dévoré des livres pour mes préparations.... L'enjeu n'est pas tant le contenu, les formules sont les mêmes dans toutes les références, que la manière de présenter. Il faut agencer le discours afin de pouvoir soutenir l'attention des étudiants sur une durée relativement longue. Ce point est certainement l'un des aspects cruciaux de l'enseignement « en live » de nos jours, que ce soit en présentiel ou en distanciel. J'ai cherché, et j'ai fini par découvrir l'ouvrage « [Econométrie](#) » de Régis Bourbonnais (Dunod, 2<sup>ème</sup> édition, 1998). J'ai tout de suite flashé sur l'organisation du discours : concept, formules, exemples illustratifs. Avec le recul aujourd'hui, je me rends compte que ce livre m'a inspiré par la suite dans ma manière de présenter les techniques quantitatives en général, y compris les algorithmes de machine learning. Quelle que soit la méthode à décrire, il faut toujours expliquer l'intérêt de ce que l'on fait, donner les points de repères importants, il faut ensuite trouver la bonne tonalité pour la rendre attrayante, en s'appuyant sur des exemples détaillés, plusieurs s'il le faut. Multiplier les cas d'usage permet de mieux cerner les tenants et aboutissants. Cela est d'autant plus vrai pour l'économétrie qui peut se révéler rapidement rébarbative lorsque l'on commence à aligner les formules matricielles.

C'est avec un très grand plaisir aujourd'hui que je rédige cette fiche de lecture consacrée à la 10<sup>ème</sup> édition de l'ouvrage (2018). Un site web (<https://regisbourbonnais.dauphine.fr/>) avec les fichiers des données illustratives l'accompagne. J'ai vu qu'une 11<sup>ème</sup> édition vient de sortir très récemment (mars 2021). Je ne l'ai pas eu en main. Il s'agit d'une évolution si je m'en tiens à la table des matières. Il semble se distinguer – entre autres – en adjoignant une fiche synoptique à chaque fin de chapitre pour que l'on retienne les aspects essentiels abordés.

Dans ce qui suit, je décrirai les différents sujets présentés dans l'ouvrage en adoptant une présentation par chapitre.



**Chapitre 1 – Qu'est-ce que l'économétrie ?** Oui, déjà il faudrait savoir de quoi on parle. Bien connue des économistes car elle permet de donner une teneur concrète aux théories économiques, l'économétrie – la terminologie tout du moins – est méconnue du « grand public ». Pourtant, il faut peu de choses pour la rendre affriolante. C'est toujours avec un amusement certain que je montre à mes étudiants le module « [machine learning](#) » du bien connu [Mooc COURSERA](#) : les trois premières semaines commencent avec des outils et techniques économétriques (régressions linéaires et logistiques, algèbre linéaire). Ainsi, mes étudiants provenant de filière statistique, très en demande de « machine learning » et de « [data science](#) » quand ils arrivent chez nous, se rendent compte qu'ils ont pour la plupart déjà pratiqué le domaine (en L1 ou L2, ceux qui viennent de filière stat/éco ont au moins vu la corrélation et la régression simple), à l'instar de M. Jourdain qui faisait de la prose sans le savoir. L'auteur choisit l'angle de l'économie dans ce chapitre, en précisant la notion de modèle, puis en détaillant le rôle de l'économétrie. Il aborde ensuite le concept de corrélation. La présentation, et cela sera vrai tout au long de l'ouvrage, est accompagnée d'un exercice corrigé didactique. Les données correspondantes (fichier Excel) sont accessibles sur [le site de l'auteur](#).

**Chapitre 2 – Le modèle de régression simple.** Comme son nom l'indique, ce chapitre est consacré à la régression linéaire simple. L'auteur s'appuie sur la fonction de consommation keynésienne pour expliciter les différentes composantes du modèle. Il traite l'estimation des paramètres (pente et constante) ensuite, non sans avoir passé en revue les hypothèses structurelles et stochastiques qui régissent les propriétés des estimateurs, et rendent possibles les calculs ainsi que l'inférence statistique. La présentation est somme toute assez classique. Je mettrais pour ma part en exergue l'accent porté sur la prédiction et le calcul des intervalles de prédiction, mis en valeur dans le cadre des données temporelles. La série d'exercices clôturant le chapitre permettent de mieux assimiler les très nombreuses formules associées à la technique.

**Chapitre 3 – Le modèle de régression multiple.** Un morceau coriace arrive avec ce chapitre consacré à la régression linéaire multiple. Ici aussi, la présentation est usuelle. Il est de toute manière difficile de faire original en la matière. Le discours est forcément guidé par les formules et le cadre inférentiel. Mon attention a été attirée par l'utilisation des indicatrices, en particulier pour le traitement de la saisonnalité. J'ai beaucoup apprécié également le thème de la régression récursive destinée à évaluer la stabilité de la régression. M'étant cantonné aux données exclusivement transversales dans mon enseignement, je n'abordais pas ces sujets. Ils sont d'importance lorsque nous traitons des données longitudinales. Je note dans les annexes du chapitre la description des sorties des logiciels utilisables dans le cadre de la régression linéaire multiple : Gretl, que [j'avais moi-même étudié en son temps](#) ; Eviews ; Excel, qui n'est certainement pas un logiciel d'économétrie, mais qui propose des fonctions nous permettant



d'obtenir des résultats viables. Je l'utilisais moi-même dans les parties introductives de mes travaux dirigés (TD.1a, TD.1b, TD.2).

**Chapitre 4 – Multicolinéarité et sélection du modèle optimal.** La question des variables pertinentes dans la régression se pose nécessairement dès que l'on traite de données un tant soit peu réalistes. L'ouvrage explore dans un premier temps les notions de corrélations brutes et partielles, leurs relations avec le coefficient de corrélation multiple de la régression (la racine carrée du fameux  $R^2$ , coefficient de détermination). Il passe ensuite aux techniques de détection de multicolinéarité, avant de conclure avec les approches pour sélectionner le modèle « optimal ». J'ai mis sciemment des guillemets à « optimal » parce qu'il n'y a certainement pas de solution exacte en la matière. Il faut voir ces méthodes comme pourvoyeurs d'autant de scénarios de solutions permettant d'appréhender les relations entre l'endogène et les explicatives.

**Chapitre 5 – Problèmes particuliers : la violation des hypothèses.** Nous abordons ces problèmes en maîtrise (Master 1) de mon temps. L'autocorrélation des erreurs et l'homoscédasticité affectent la variance des estimateurs ; l'erreur sur les variables induit des estimateurs biaisés. L'auteur présente différents tests qui permettent d'identifier les situations à problème. Il s'attaque aux solutions par la suite. Difficile d'échapper au déluge de formules matricielles ici. Et dire que nous devons les recopier les gribouillis de l'enseignant au tableau alors, avec toutes les erreurs que cela pouvait impliquer. Heureusement, plus personne ne fait cela aujourd'hui. On retrouve ces formules sur n'importe quel support accessible en ligne ou dans des cours sur YouTube. Les étudiants peuvent consacrer leur énergie à des choses autrement plus constructives. Pour en revenir à l'ouvrage, les exemples en abondance permettent d'appréhender concrètement les approches et la lecture des résultats.

**Chapitre 6 – Les modèles non linéaires.** Ce 6<sup>ème</sup> chapitre est consacré aux modèles non linéaires. La solution la plus simple consiste à se ramener à une configuration linéaire via des artifices de calcul (passage aux logarithmes par exemple). Lorsque cela n'est pas possible, on passe par des algorithmes itératifs dans laquelle l'équation de régression est linéarisée localement à l'aide d'un développement limité de Taylor (page 184). L'auteur s'attarde peu sur les aspects théoriques. Il s'attache surtout à montrer la démarche sur des exemples numériques résolus notamment à l'aide du logiciel Eviews.

**Chapitre 7 – Les modèles à décalages temporels.** Lorsque l'on traite de données longitudinales, les effets sont rarement synchrones. Ce sont les revenus de la période précédente qui pèsent sur la consommation d'aujourd'hui, les efforts promotionnels pour un produit induit des ventes supplémentaires dans le futur. Les modèles à décalages peuvent inclure comme variables retardées aussi bien la variable endogène que les variables exogènes (page 192).



L'auteur présente les modèles auto-régressifs (section 1), puis les modèles à retards échelonnés (section 2). La question de l'horizon du retard est posée. Pour éviter le sur-apprentissage et l'obtention des coefficients totalement erratiques, des contraintes peuvent être appliquées pour que les coefficients successifs prennent une certaine forme (ex. une décroissance polynomiale). A bien y regarder, on peut y voir une forme de régularisation en réalité. Franchement, maintenant que j'ai un recul assez conséquent sur les cultures économétriques et de machine learning, je trouve qu'ils avaient des idées assez épatantes ces économètres. Il faut un peu s'accrocher pour suivre les exemples dans ce chapitre. Mais le jeu en vaut la chandelle si on veut pouvoir reproduire les approches sur d'autres données.

**Chapitre 8 – Introduction aux modèles à équations simultanées.** On rentre de plain-pied dans la modélisation économétrique dans ce chapitre. Un système économique se résume souvent par un système d'équations mathématiques où les rôles des variables sont enchevêtrés (parfois endogènes, parfois exogènes, avec ou pas des retards). On parle de système d'équations structurelles. Deux problèmes se posent alors : est-ce que l'architecture du modèle permet l'estimation des paramètres, on parle d'identification ; quelles sont les techniques que l'on peut mettre en œuvre pour estimer les coefficients. L'affaire n'est pas triviale, c'est le moins qu'on puisse dire. L'exemple du modèle de Klein (page 249) est très représentatif des difficultés que l'on rencontre lorsque nous sommes confrontés à ce type d'analyse. Ici aussi, il faut vraiment s'accrocher aux branches pour suivre le détail des étapes.

**Chapitre 9 – Éléments d'analyse des séries temporelles.** L'auteur traite des séries chronologiques dans ce chapitre. On remarquera que la majorité des techniques vues précédemment pouvaient s'appliquer aux données longitudinales. C'est la spécificité des techniques qui fait l'originalité de cette partie. J'y trouve toute une série de termes et de noms de méthodes qui me rappellent de vieux souvenirs : stationnarité, test de bruit blanc, test de racine unitaire (Dickey-Fuller, etc.). Ouh là là, la nostalgie me guette. La 3<sup>ème</sup> section est consacrée aux processus ARIMA, la 4<sup>ème</sup> à la méthode de Box-Jenkins destinée à l'estimation des paramètres. L'exposé reste succinct quand-même. L'auteur rappelle incidemment qu'il a rédigé un livre spécialisé concernant le traitement des séries chronologiques (R. Bourbonnais et M. Terraza, « Analyse des séries temporelles : applications à l'économie et à la gestion », Dunod, 2016).

**Chapitre 10 – La modélisation VAR.** Le modèle VAR est une généralisation multidimensionnelle du processus AR (autorégressive), sans la partie MA donc (moving-average). Les questions usuelles sont traitées : estimation, détermination des retards. Je retiendrai pour ma part les sections 3 (« Dynamique d'un modèle VAR ») et 4 (« La causalité ») qui placent la technique au centre des préoccupations de la modélisation économique. Une approche statistique n'a de sens



que si elle permet de répondre à des problématiques concrètes d'un domaine d'étude. C'est la fameuse dimension « métier » de la data science. Dans le cadre de l'économétrie, la finalité économique des analyses sont prégnantes, les exemples illustratifs mises en avant par l'auteur sont assez éloquentes.

**Chapitre 11 – La cointégration et le modèle à correction d'erreur.** Il s'agit d'établir une double relation dans les modèles à correction d'erreur : une tendance commune (la relation de cointégration), et la liaison réelle entre les variables (stationnaires). Dans la relation à deux variables, l'estimation à l'aide des moindres carrés ordinaires (MCO) de deux régressions simples est très facile à mettre en œuvre. L'affaire se corse lorsque nous souhaitons traiter  $k$  variables (section 4). L'auteur donne un aperçu des étapes, agrémenté d'un exemple. Mais il s'agit plutôt d'un survol destiné à nous mettre le pied à l'étrier. Il en faudra plus pour approfondir le domaine.

**Chapitre 12 – Introduction à l'économétrie des variables qualitatives.** **Le traitement des variables endogènes qualitatives, nominales ou ordinales**, me fascinait quand j'étais étudiant. La plupart des références en économétrie se contentent d'une introduction succincte tassée en fin d'ouvrage souvent. C'est aussi un peu le cas ici. On voit quand-même les enjeux de l'approche. Les exemples numériques participent heureusement à la bonne perception des méthodes. Je mettrai une mention spéciale au modèle TOBIT pour l'explication et la prédiction des variables limitées (censurées). Ses applications sont nombreuses et concrètes. Pourtant la méthode est très souvent absente des ouvrages d'obédience « machine learning ».

**Chapitre 13 – Introduction à l'économétrie des données de panel.** Le chapitre consacré aux données de panel clôture l'ouvrage. Pour rappel, il s'agit de données composées d'une double dimension : transversale, les individus sont caractérisés par plusieurs variables ; longitudinale, les individus sont observés dans le temps. Lorsque ce sont exactement les mêmes individus, on parle de cohorte. Dans d'autres contextes, on parlerait de « données évolutives ». Il s'agit d'une introduction ici. La présentation est concise, s'attachant surtout à mettre en avant les approches utilisées (les contraintes que nous nous imposons, ou les degrés de liberté que nous nous autorisons) selon les informations que l'on souhaite exploiter dans les données. Je ne voudrais pas être trop répétitif, mais ce type problématique est souvent absente de la littérature « machine learning », pourtant les applications concrètes sont nombreuses.

**En conclusion de cette fiche de lecture**, je dirais qu'en matière de support pédagogique, le fond est important, personne ne dira le contraire, mais la forme l'est tout autant, surtout quand il s'agit d'éveiller (de titiller) les ressorts de compréhension dans l'esprit des apprenants. Tout le monde l'aura compris, la lecture de l'ouvrage de Régis Bourbonnais m'a beaucoup plu, parce que j'y ai justement trouvé une démarche qui me paraît judicieuse pour l'enseignement de



l'économétrie, pour l'enseignement des techniques quantitatives en général. La pratique ne peut pas être dissociée de la théorie. J'ai trouvé très intéressante l'utilisation qui est faite des exemples didactiques tout au long du livre.