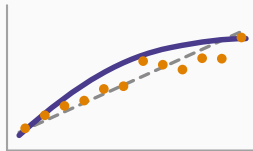


Économétrie avancée

Chapitre 16 : Modèles à équations simultanées



Jaouad Madkour

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

Avant de commencer

La nature des modèles à équations simultanées

Le biais de simultanéité

Identification et estimation

SEM avec séries temporelles et données de panel

Synthèse

Avant de commencer

Pourquoi ce chapitre ?

Variable omise et erreur de mesure (chapitre 9) ne sont pas les seules sources d'endogénéité. Quand y et un régresseur sont **déterminés conjointement** — typiquement par un mécanisme d'équilibre offre-demande — les MCO échouent structurellement. Ce chapitre montre comment reconnaître un système d'équations simultanées valide, et comment l'estimer par variables instrumentales.

Rappel express : variables instrumentales et 2SLS

Si x est corrélé à u dans $y = \beta_0 + \beta_1 x + u$, une variable instrumentale z — corrélée à x mais non à u — permet d'identifier β_1 . Les **doubles moindres carrés (2SLS)** appliquent la procédure en deux étapes : (i) régresser x sur z et tous les régresseurs exogènes, retenir les valeurs ajustées $\hat{x}$; (ii) régresser y sur $\hat{x}$ et les régresseurs exogènes. Ce chapitre applique directement cet outil; le traitement complet des variables instrumentales fera l'objet d'un chapitre ultérieur.

La nature des modèles à équations simultanées

Équation structurelle et interprétation *ceteris paribus*

Un **modèle à équations simultanées (SEM)** rassemble plusieurs équations dont **chacune** doit avoir une interprétation causale autonome — typiquement, le comportement de deux agents économiques distincts (offre des travailleurs vs demande des employeurs). Les variables déterminées à l'intérieur du système sont **endogènes**; celles déterminées à l'extérieur sont **exogènes**.

Offre et demande de travail agricole

$$h_t^s = \alpha_1 w_t + \beta_1 z_{1t} + u_{1t} \quad (\text{offre}), \quad h_t^d = \alpha_2 w_t + \beta_2 z_{2t} + u_{2t} \quad (\text{demande}),$$

avec la condition d'équilibre $h_t^s = h_t^d = h_t$. z_1 (salaire manufacturier alternatif) et z_2 (surface agricole) sont des **variables observées** propres à chaque équation — c'est précisément ce qui permet de distinguer laquelle est l'offre et laquelle est la demande, et, comme on le verra, de les estimer séparément.

Un critère simple pour reconnaître un bon SEM

Demander « que se passerait-il si y_2 changeait de façon exogène, toutes choses égales par ailleurs ? » doit avoir un sens économique pour **chaque** équation. Un contre-exemple classique : modéliser dépenses de logement et épargne d'un même ménage comme un système simultané n'a pas de sens, car les deux décisions résultent du **même** problème de maximisation d'utilité — il n'existe pas de variable exogène propre à l'une qui n'affecte pas l'autre.

Le biais de simultanéité

Pourquoi les MCO échouent

Corrélation entre régresseur et erreur

Dans le système $y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$, $y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2$ (avec $\alpha_1 \alpha_2 \neq 1$), la **forme réduite** exprime y_2 en fonction des seules variables exogènes :

$$y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2, \quad v_2 = \frac{\alpha_2 u_1 + u_2}{1 - \alpha_1 \alpha_2}.$$

Puisque v_2 est fonction de u_1 , y_2 est en général **corrélé** à u_1 – sauf cas particulier ($\alpha_2 = 0$ et u_1, u_2 non corrélées). C'est le **biais de simultanéité**.

Application : police et criminalité

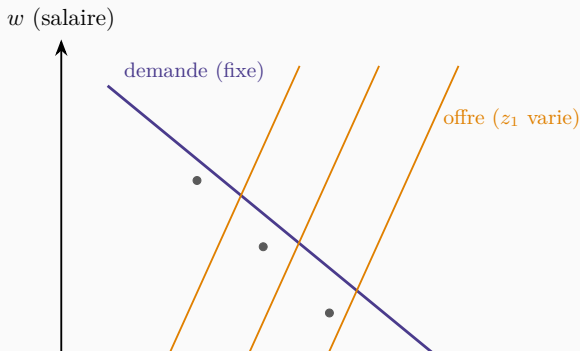
Estimer $murdpc = \alpha_1 polpc + \dots + u_1$ par MCO surestime généralement la corrélation positive : les villes réagissent à une criminalité élevée en augmentant leurs effectifs de police ($polpc = \alpha_2 murdpc + \dots + u_2$, $\alpha_2 > 0$), ce qui contamine la relation observée. Le signe du biais se déduit de $\text{Cov}(y_2, u_1) \propto \alpha_2 / (1 - \alpha_1 \alpha_2)$: avec $\alpha_2 > 0$, l'effet dissuasif d'un renforcement policier est **sous-estimé** (biais vers zéro ou de signe opposé) par les MCO.

Identification et estimation

La condition de rang

Conditions d'ordre et de rang

Pour le système $y_1 = \beta_{10} + \alpha_1 y_2 + \mathbf{z}_1 \beta_1 + u_1$, $y_2 = \beta_{20} + \alpha_2 y_1 + \mathbf{z}_2 \beta_2 + u_2$: la première équation est **identifiée** si et seulement si la seconde équation contient au moins une variable exogène (à coefficient non nul) **exclue** de la première. La **condition d'ordre** (au moins une variable exclue) est nécessaire; la **condition de rang** (cette variable a un coefficient réellement non nul dans l'autre équation, testable par t ou F) est nécessaire et suffisante.



Équations sur-identifiées, juste identifiées, non identifiées

La condition d'ordre se généralise : une équation est identifiée seulement si le nombre de variables exogènes **exclues** est au moins égal au nombre de variables endogènes à droite. Une équation est **sur-identifiée** si elle dispose de plus d'instruments que nécessaire (testable par le test de sur-identification), **juste identifiée** si le nombre correspond exactement, et **non identifiée** si aucune variable exogène n'est exclue.

SEM avec séries temporelles et données de panel

Un cas particulier utile : les retards comme instruments naturels

Une valeur **retardée** d'une variable endogène (Y_{t-1}) peut être traitée comme **prédéterminée** (donc utilisable comme instrument) si l'erreur structurelle courante lui est non corrélée — hypothèse plausible quand l'information disponible à $t - 1$ ne peut être affectée par un choc survenant en t . C'est le principe exploité pour tester l'hypothèse du revenu permanent : les erreurs sont supposées non corrélées à toute l'information disponible à $t - 1$, ce qui rend n'importe quelle variable retardée valide comme instrument.

SEM à effets fixes

Pour $y_{1it} = \alpha_1 y_{2it} + \mathbf{z}_{1it} \beta_1 + a_i + u_{1it}$, différencier élimine a_i :

$\Delta y_{1it} = \alpha_1 \Delta y_{2it} + \Delta \mathbf{z}_{1it} \beta_1 + \Delta u_{1it}$. Il faut alors un instrument pour Δy_{2it} – nécessairement une variable **qui varie dans le temps**, condition parfois difficile à satisfaire (l'expérience professionnelle, par exemple, augmente d'exactly une unité chaque année pour tout individu en emploi continu, donc $\Delta \text{exper}_{it} = 1$ pour tous : elle ne peut servir d'instrument différencié).

Application : prisons et criminalité (Levitt, 1996)

Pour estimer l'effet causal de la population carcérale sur la criminalité, Levitt utilise des variables indicatrices de **litiges de surpopulation carcérale** comme instruments pour $\Delta \log(\text{prison})$ – un choc plausiblement exogène aux fluctuations de criminalité. L'estimation 2SLS ($\hat{\alpha} \approx -1,02$) diffère fortement de l'estimation MCO groupée ($\hat{\alpha} \approx -.18$), illustrant l'ampleur du biais de simultanéité dans la relation prison-criminalité brute.

Synthèse

L'essentiel du chapitre 15

- Un SEM valide exige que **chaque** équation ait une interprétation causale autonome — le cas archétype étant offre et demande.
- La simultanéité corréle un régresseur endogène à l'erreur, biaisant et rendant incohérents les MCO; direction du biais généralement liée au signe du coefficient de rétroaction.
- Identification : au moins une variable exogène exclue de l'équation d'intérêt, avec coefficient réellement non nul dans l'autre équation (condition d'ordre + condition de rang); l'équation identifiée s'estime alors par 2SLS.
- Séries temporelles et panels permettent des instruments naturels (retards, différenciation + instruments variant dans le temps), au prix d'hypothèses supplémentaires sur l'exogénéité stricte des variables prédéterminées.

Vers le chapitre 16

Ce chapitre a traité l'endogénéité provenant d'un système d'équations. Le chapitre suivant aborde un autre type de complication : les variables dépendantes **limitées** — binaires, censurées, ou observées seulement pour un sous-échantillon non aléatoire — qui exigent des modèles au-delà de la régression linéaire classique.