

Fiche de révision pour l'économétrie non linéaire

Tran Viet Chi

Ce document est un support de TD, qui s'inspire du cours fait par B. Salanié (<http://www.crest.fr/pageperso/lei/salanie/salanie.htm>).

On considère un modèle paramétrique ou semi-paramétrique, le "vrai" paramètre θ_0 appartenant à Θ . Notre but est d'estimer ce paramètre à l'aide de n observations statistiques. Pour fixer les idées, les rappels de cette fiche s'appliquent au cas de la modélisation de la loi d'une variable y en présence d'explicatives x . On observe n couples i.i.d. de variables $(y_i, x_i)_{1 \leq i \leq n}$. Bien sûr, les lois des y_i conditionnellement aux x_i peuvent, elles, être différentes!

1 M-estimateurs

1.1 Définitions et exemples

Les M-estimateurs sont fondés sur la maximisation d'un critère :

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} Q_n(\theta). \quad (1)$$

Q_n est le critère qui dépend des n observations et de θ .

Par exemple :

1. Si $Q_n(\theta) = (1/n) \sum_{i=1}^n \log f(y_i, x_i, \theta)$ où f est la densité de la loi de (y, x) par rapport à une mesure de référence, alors, on obtient l'estimateur du maximum de vraisemblance.
2. Si $Q_n(\theta) = -(1/n) \sum_{i=1}^n (y_i - m(x_i, \theta))'(y_i - m(x_i, \theta))$ où $m(x, \theta) = E(y | x, \theta)$ alors on obtient l'estimateur des moindres carrés non linéaires (MCNL, régression analogue des MCO mais non linéaire),
3. Si $Q_n(\theta) = -\|(1/n) \sum_{i=1}^n h(y_i, x_i, \theta)\|_{S_n}^2$ où h est une fonction liée au problème, alors on obtient un estimateur des moments généralisés (GMM). La maximisation correspond alors à la minimisation du pendant empirique d'une relation estimante théorique :

$$E(h(y, x, \theta_0)) = 0.$$

Les exemples ci-dessus définissent des types d'estimateurs (EMV, MCNL, GMM...) qui ne sont souvent pas exclusifs (un estimateur peut être à la fois un estimateur GMM et MCNL etc.).

1.2 Critères de convergence

Proposition 1.1. Convergence presque-sûre

Si :

1. les fonctions Q_n sont continues,
2. la convergence de la suite de fonctions Q_n est uniforme en θ presque sûrement,
3. la limite $Q_\infty(\theta)$ a un maximiseur unique qui est θ_0

alors $(\hat{\theta}_n)_{n \in \mathbb{N}}$ converge presque sûrement vers θ_0 .

Les hypothèses de la proposition précédente ne sont parfois pas facilement vérifiables (par exemple la convergence uniforme).

Proposition 1.2. Convergence en probabilité

Si on a la propriété de lipschitzianité suivante :

$$\forall \theta, \theta' \in \Theta, |Q_n(\theta) - Q_n(\theta')| \leq B_n \|\theta - \theta'\|,$$

avec B_n une suite bornée en probabilité, alors la suite $(\hat{\theta}_n)_{n \in \mathbb{N}}$ est convergente en probabilité.

Pour les théorèmes de normalité asymptotique, on utilisera souvent soit des résultats liés aux méthodes utilisées (MCNL, GMM...), soit le TCL et la δ -méthode.

On utilise le théorème de Slutsky pour remplacer les termes inconnus (par exemple les matrices de produits scalaires inconnues) par des estimations.

2 EMV

Il faut revoir le cours sur l'EMV. Rappelez-vous que souvent, l'EMV est convergent et efficace. Cependant, il peut-être difficile à calculer, et est peu robuste (si on se trompe de lois, on obtient un mauvais estimateur) c'est pourquoi on peut avoir recours à d'autres méthodes d'estimation.

3 GMM

3.1 Principe

On peut mettre en oeuvre les GMM lorsque l'on dispose d'une relation estimante de la forme:

$$E(h(y, x, \theta_0)) = 0. \tag{2}$$

Le principe est de remplacer l'espérance par une quantité empirique : $(1/n) \sum_{i=1}^n h(y_i, x_i, \theta)$. Il est possible que cette quantité ne s'annule pour aucun θ ; au lieu d'annuler, on minimise alors la norme :

$$\hat{\theta}_{GMM} = \arg \min_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \theta) \right\|_{S_n}^2 = \arg \min_{\theta \in \Theta} \left(\frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \theta) \right)' S_n \left(\frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \theta) \right). \tag{3}$$

3.2 Convergence et normalité asymptotique

Proposition 3.1. Convergence

Si :

1. h est continue,
2. $(1/n) \sum_{i=1}^n h(y_i, x_i, \theta)$ converge vers $E(h(y, x, \theta))$ uniformément en θ ,
3. S_n converge vers S_∞ définie positive,
4. le modèle est asymptotiquement identifiable au sens où :

$$[E(h(y, x, \theta))' S_\infty E(h(y, x, \theta)) = 0] \Rightarrow [\theta = \theta_0].$$

Alors $\hat{\theta}_{GMM}$ converge presque sûrement vers θ_0 .

Proposition 3.2. Variance asymptotique et lois asymptotiques

i) Le choix optimal pour S_n est obtenu lorsque $S_n \rightarrow S_\infty = (\text{Var}(h(y, x, \theta_0)))^{-1}$ lorsque $n \rightarrow \infty$, la convergence étant en probabilité.

ii) Dans ce cas, la matrice de variance asymptotique de $\hat{\theta}_{GMM}$ est:

$$\left\{ E \left(\frac{\partial h}{\partial \theta'}(y, x, \theta_0) S_\infty \frac{\partial h'}{\partial \theta}(y, x, \theta_0) \right) \right\}^{-1}.$$

iii) Pour le choix précédent de S_n , on a :

$$\xi_n := n \left\| \frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \hat{\theta}_{GMM}) \right\|_{S_n}^2 \rightarrow_{n \rightarrow \infty} \chi^2(\text{nb rel estimantes} - \text{dim } \theta),$$

la convergence étant une convergence en loi. Cette dernière relation est utile pour tester $H_0 : E(h(y, x, \theta)) = 0$. La région de rejet pour le seuil α est alors :

$$W = \{ \xi_n \geq q_{1-\alpha}(\chi^2(\text{nb rel estimantes} - \text{dim } \theta)) \}.$$

On peut noter que le choix de S_n convergeant vers S_∞ (point *i*) de la Proposition 3.2) n'est pas unique. Ces choix donnent la même matrice de variance-covariance asymptotique (point *ii*) de la Proposition 3.2).

3.3 Mise en oeuvre

Méthode de calcul en deux étapes : on réalise une première étape préalable pour estimer la matrice S_n qui donnerait un estimateur optimal au sens de la variance asymptotique. La Proposition 3.2 nous dit comment faire ce choix.

1. Dans une première étape, on calcule un estimateur $\tilde{\theta}_n$ avec S_n quelconque, généralement la matrice identité :

$$\tilde{\theta}_n = \arg \min_{\theta \in \Theta} \left(\frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \theta) \right)' \left(\frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \theta) \right).$$

Cet estimateur est convergent, mais pas optimal (le choix de $S_n = Id$ ne correspond pas au choix optimal dicté par la Proposition 3.2).

2. On définit alors :

$$S_n = \left(\frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \tilde{\theta}_n) h(y_i, x_i, \tilde{\theta}_n)' \right)^{-1},$$

qui converge vers $S_\infty = (Var(h(y, x, \theta_0)))^{-1}$. On réestime alors avec ce choix de S_n :

$$\hat{\theta}_{GMM} = \arg \min_{\theta \in \Theta} \left(\frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \theta) \right)' S_n \left(\frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \theta) \right).$$

D'après la Proposition 3.2, l'estimateur $\hat{\theta}_{GMM}$ est asymptotiquement optimal.

Méthode de calcul en une étape

Elle revient à choisir :

$$S_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \theta) h(y_i, x_i, \theta)' \right)^{-1},$$

et à minimiser

$$\left(\frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \theta) \right)' S_n(\theta) \left(\frac{1}{n} \sum_{i=1}^n h(y_i, x_i, \theta) \right).$$

Dans la méthode en deux étapes, on "gelait" la matrice de produit scalaire avant chaque minimisation, ce qui permettait de minimiser des fonctions quadratiques. Ici, on ne fait qu'une minimisation, mais comme $S_n(\theta)$ dépend de θ , la fonction à minimiser n'est plus nécessairement quadratique, et cela peut-être plus difficile numériquement.

4 MCNL

4.1 Principe

Dans les modèles de régression, y est une fonction bruitée de x :

$$y = m(x, \theta) + \varepsilon,$$

où $m(x, \theta) = E(y | x, \theta)$.

La régression non-linéaire, tout comme pour la régression linéaire, minimise la variance des résidus :

$$\hat{\theta}_{MCNL} = \arg \min_{\theta \in \Theta} \left\| \begin{pmatrix} y_1 - m(x_1, \theta) \\ \vdots \\ y_n - m(x_n, \theta) \end{pmatrix} \right\|_{S_n}^2. \quad (4)$$

Lorsque $S_n = Id_n$ la matrice identité de \mathbb{R}^n , on obtient les MCNL.

Il n'y a *a priori* pas de raison de choisir la norme euclidienne pour calculer l'écart entre le vecteur des observations $(y_1, \dots, y_n)'$ et le vecteur des régresseurs $(m(x_1, \theta), \dots, m(x_n, \theta))'$ et on peut chercher à minimiser la norme calculée dans une autre métrique S_n . On obtient alors les MCNLQG (quasi-généralisés).

Lorsque la matrice S_n est diagonale par blocs de la forme :

$$S_n = \begin{pmatrix} a_n(x_1) & & 0 \\ & \ddots & \\ 0 & & a_n(x_n) \end{pmatrix}, \quad (5)$$

alors :

$$\hat{\theta}_{MCNLQG} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n (y_i - m(x_i, \theta))' a_n(x_i) (y_i - m(x_i, \theta)). \quad (6)$$

C'est ce qu'on supposera dans la suite.

4.2 Propriétés

1. L'estimateur $\hat{\theta}_n$ est convergent quel que soit la matrice a_n dans (5).
2. La variance asymptotique de $\hat{\theta}_n$ est :

$$Var(a_n(x)) = Var \left(B(a_n)^{-1} a_n(x)^{-1} \frac{\partial m'}{\partial \theta}(x, \theta_0) (y - m(x, \theta_0)) \right),$$

avec :

$$B(a_n) = E \left(a_n(x)^{-1} \frac{\partial m'}{\partial \theta}(x, \theta_0) \frac{\partial m}{\partial \theta'}(x, \theta_0) \right).$$

3. Le choix optimal de a_n est tel que $a_n \rightarrow a^*$ en probabilité, lorsque $n \rightarrow \infty$, avec $a^*(x) = (Var(y | x, \theta_0))^{-1}$.
4. Pour le choix de S_n précédent, l'estimateur des MCNLQG est convergent et asymptotiquement efficace.

4.3 Mise en oeuvre

La mise en oeuvre des MCNLQG est classique et se fait en 2 étapes. Le principe est le même que pour les MCQG :

1. On commence par calculer l'estimateur MCNL non pondéré décrit au paragraphe 4.1. On obtient un estimateur convergent $\tilde{\theta}_n$ de θ_0 (mais non optimal) :

$$\tilde{\theta}_n = \arg \min_{\theta \in \Theta} \sum_{i=1}^n (y_i - m(x_i, \theta))' (y_i - m(x_i, \theta)).$$

Ce paramètre $\tilde{\theta}_n$ nous permet d'estimer l'inverse de la variance conditionnelle des y , $S_n(x, \tilde{\theta}_n) = (V(y | x, \tilde{\theta}_n))^{-1}$.

2. Dans un second temps, on minimise:

$$\hat{\theta}_{MCQG} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n (y_i - m(x_i, \theta))' S_n(x_i, \tilde{\theta}_n) (y_i - m(x_i, \theta)).$$

Ainsi, on ne minimise que des fonctions quadratiques, d'où un gain en procédures numériques.

5 MCA

5.1 Principe

Les MCA sont des estimateurs de distance minimale. Le principe est le suivant :

1. On dispose de relations entre le paramètre θ_0 et des paramètres auxiliaires π_0 : $g(\pi_0, \theta_0) = 0$, et telles que $g(\pi_0, \theta) = 0 \Rightarrow \theta = \theta_0$.
2. On dispose d'estimateurs $\hat{\pi}_n$ convergents et asymptotiquement normaux pour π_0 :

$$\sqrt{n} (\hat{\pi}_n - \pi_0) \rightarrow_{n \rightarrow \infty} \mathcal{N}(0, \Omega_0),$$

la convergence étant une convergence en loi.

3. L'estimateur MCA est alors :

$$\hat{\theta}_{MCA} = \arg \min_{\theta \in \Theta} \|g(\hat{\pi}_n, \theta)\|_{S_n}^2 = \arg \min_{\theta \in \Theta} (g(\hat{\pi}_n, \theta))' S_n (g(\hat{\pi}_n, \theta)). \quad (7)$$

L'interprétation heuristique est la suivante : lorsque g s'interprète comme une distance, on choisit $\hat{\theta}_{MCA}$ proche de $\tilde{\pi}_n$, lui-même proche de π_0 si n est assez grand. Comme π_0 est à distance 0 de θ_0 , $\hat{\theta}_{MCA}$ sera également un estimateur convergent de θ_0 .

5.2 Convergence et lois asymptotiques

Proposition 5.1. Convergence

Si :

1. g est continue,
2. S_n converge vers S_∞ définie positive,
3. on a la propriété d'identifiabilité suivante :

$$[g(\pi_0, \theta)' S_\infty g(\pi_0, \theta) = 0] \Rightarrow [\theta = \theta_0],$$

alors $\hat{\theta}_{MCA}$ est convergent.

Proposition 5.2. Variance asymptotique et lois asymptotiques

i) Pour le choix de :

$$S_n \rightarrow S_\infty = \left(\frac{\partial g}{\partial \pi'}(\pi_0, \theta_0) \Omega_0 \frac{\partial g'}{\partial \pi}(\pi_0, \theta_0) \right)^{-1},$$

où Ω_0 est la variance asymptotique des $\hat{\pi}_n$, on a un estimateur optimal de variance asymptotique :

$$\left(\frac{\partial g}{\partial \pi'}(\pi_0, \theta_0) S_\infty \frac{\partial g'}{\partial \pi}(\pi_0, \theta_0) \right)^{-1}.$$

ii) Il est possible de tester $H_0 : g(\pi_0, \theta_0) = 0$. Sous l'hypothèse nulle et pour le choix de S_n précédent :

$$n g(\hat{\pi}_n, \hat{\theta}_{MCA})' S_n g(\hat{\pi}_n, \hat{\theta}_{MCA}) \rightarrow_{n \rightarrow \infty} \chi^2(\dim \text{relations} - \dim \theta),$$

la convergence étant une convergence en loi.

5.3 Mise en oeuvre

Méthode de calcul en deux étapes

1. Dans une première étape, on calcule un estimateur $\tilde{\theta}_n$ avec S_n quelconque, généralement la matrice identité :

$$\tilde{\theta}_n = \arg \min_{\theta \in \Theta} (g(\hat{\pi}_n, \theta))' (g(\hat{\pi}_n, \theta)).$$

Cet estimateur est convergent, mais pas optimal.

2. On définit alors :

$$S_n = \left(\frac{\partial g}{\partial \pi'}(\hat{\pi}_n, \tilde{\theta}_n) \hat{\Omega} \frac{\partial g'}{\partial \pi}(\hat{\pi}_n, \tilde{\theta}_n) \right)^{-1},$$

où $\hat{\Omega}$ est une estimation de Ω_0 la matrice de variance-covariance asymptotique de $\hat{\pi}_n$, et on réestime avec ce choix de S_n :

$$\hat{\theta}_{MCA} = \arg \min_{\theta \in \Theta} (g(\hat{\pi}_n, \theta))' S_n (g(\hat{\pi}_n, \theta)).$$

6 Pseudo-vraisemblance

Le principe du pseudo-maximum de vraisemblance est de remplacer la vraisemblance du modèle, qui peut-être compliquée, par la vraisemblance d'une loi plus simple de même espérance (on parle alors de pseudo-maximum de vraisemblance d'ordre 1) ou de même espérance et même variance (PMV d'ordre 2) etc.

$$\hat{\theta}_{PMV} = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \log g(x_i, \theta), \quad (8)$$

où g n'est pas la densité du modèle, mais celle d'une autre loi bien choisie (mêmes premiers moments que la loi d'intérêt par exemple).

1. L'estimateur PMV à l'ordre 1, $\hat{\theta}_{PMV}$, est convergent lorsque le pseudo-modèle est exponentiel d'espérance m , c'est à dire lorsque la log-vraisemblance d'une observation dans le pseudo-modèle est de la forme :

$$\log g(x, m) = A(m) + B(x) + C(m)x.$$