

LE PSEUDO-MAXIMUM DE VRAISEMBLANCE

THEORIQUE ET SIMULE

Application au cas des modèles de déséquilibre

BEN JELILI Riadh*

HAMOUCHE Salima**

MIHOUBI Ferhat***

* CREST, Université de Bourgogne (LATEC) et Université de Paris II.

** CREST et Université du Maine.

*** CREST, Université de Paris I (CME) et Ministère du Travail.

Nous tenons à remercier Christian GOURIEROUX et Pierre MORIN pour leurs remarques et leurs suggestions. Toutefois, nous sommes les seuls responsables des éventuelles erreurs.

Sommaire

0- Introduction

I- Notations et définitions

II- présentation générale du PMV

II.1- Information discriminante de Kullback convergence du PMV

II.2- Conditions nécessaires et suffisantes de convergence

III- Présentation de la version simulée du PMV

II.1- Bref rappel historique des méthodes simulées

II.2- Le pseudo-maximum de vraisemblance simulé

IV- Application au cas des modèles de déséquilibre

IV.1- Présentation du modèle et des approches alternatives d'estimation

IV.2- Estimation par le PMV et par le PMVS

IV.3- Application à l'emploi en déséquilibre et en présence de variables latentes retardées

Annexe 1

Annexe 2

Références bibliographiques

0- INTRODUCTION

Dans certains cas, les modèles économétriques conduisent à des expressions complexes de la distribution conditionnelle des variables endogènes. Les fonctions de vraisemblance qui leur sont associées sont difficilement manipulables, voire intraitables : Il devient difficile d'estimer les paramètres d'intérêt de ces modèles. Les exemples de telles situations sont nombreux et couvrent plusieurs domaines de l'économie : les modèles de choix discrets avec plusieurs alternatives, les modèles de déséquilibre en présence d'une dynamique sur les variables latentes, les modèles de recherche d'emploi, les modèles non-linéaires avec des coefficients aléatoires, les modèles non-linéaires avec anticipation rationnelle, les modèles en temps continu avec volatilité stochastique, etc. Pour contourner ces difficultés, on peut approximer la fonction de vraisemblance du modèle considéré, ou bien approximer le modèle lui-même et maximiser la fonction de vraisemblance qui lui est associée¹. Cette dernière solution est adoptée par la procédure du Pseudo-Maximum de Vraisemblance (PMV).

Les méthodes semi-paramétriques du PMV développées par White (1982), Burguete, Gallant et Suza (1982) et Gouriéroux, Monfort et Trognon (1984), combinées avec celles du PMV Simulé (PMVS) présentées par Laroque et Salanié (1989,1992), permettent d'apporter une plus grande souplesse dans la représentation de la réalité. La solution adoptée consiste alors à ne pas spécifier complètement la structure paramétrique du modèle considéré, ce qui justifie le qualificatif "*semi-paramétrique*"² associé à ces méthodes d'estimation. Ces méthodes reposent sur l'information contenue dans les moments du premier et du second ordre de la variable endogène. Pourtant, dans certains cas, les expressions analytiques de ces moments ne sont pas explicites. Le recours à des simulations stochastiques permet alors une approximation empirique de ceux-ci, tout en conservant les bonnes propriétés de convergence et d'efficacité des estimateurs ainsi obtenus. La maîtrise grandissante des procédures d'estimation par simulation élargit en conséquence le domaine d'application du PMV.

Cette note méthodologique a pour but d'explicitier, à travers un exemple concret, la mise en oeuvre de la procédure d'estimation par le PMV et par le PMV

¹Voir à ce sujet le numéro spécial de Journal of Applied Econometrics, Volume 8, Décembre 1993.

²Méthode qui ne porte pas nécessairement sur l'ensemble des paramètres régissant les observations.

Simulé³. Les modèles de déséquilibre, par essence non linéaires, permettent en effet d'illustrer l'utilité de cette procédure.

Après avoir introduit quelques éléments de définitions (section 1), nous présenterons, dans une deuxième section, le principe général du PMV. La version simulée du PMV sera ensuite examinée (section 3). La dernière partie concerne l'adaptation de cette procédure d'estimation aux modèles de déséquilibre, en général, et à un modèle de déséquilibre sur le marché du travail en particulier.

I- NOTATIONS ET DEFINITIONS

Pseudo-modèle et identifiabilité au premier et au second ordre

D'une façon générale, un modèle économétrique peut être exprimé sous la forme :

$$y_t = \phi(x_t, \theta, u_t)$$

On souhaite expliquer pour chaque observation t , $t = 1, \dots, T$, les valeurs prises par un vecteur y_t , de dimension G , de variables endogènes observables en fonction, d'un vecteur x_t de variables exogènes, de dimension K , et d'un vecteur d'erreur u_t , de dimension G :

Il est parfois plus facile de calculer les premiers moments de la variable endogène observable y_t , que d'inverser la forme réduite du modèle⁴. Dans ce cas, la liaison entre y_t et x_t n'est modélisée que par l'intermédiaire de la seule moyenne conditionnelle de y_t sachant x_t . On définit ainsi le **pseudo-modèle** suivant :

$$y_t = m(x_t, \theta) + u_t, \text{ avec } E(y_t / x_t, \theta) = m(x_t, \theta).$$

Le pseudo-modèle est dit spécifié au premier ordre si la liaison entre y_t et x_t n'est décrite que par l'espérance conditionnelle $m(x_t, \theta)$. La connaissance de θ rend donc complètement compte de l'espérance de y_t sachant x_t .

θ est dit identifiable au premier ordre si pour toute valeur θ_0 :

³On ne s'attardera pas sur les détails techniques qui ont été présentés extensivement par ailleurs (voir les références bibliographiques)

⁴L'estimation de ce modèle par le maximum de vraisemblance (MV) nécessite la réécriture du modèle en exprimant les résidus en fonction des variables (y_t, x_t) et des paramètres θ .

$$m(x_t, \theta) = m(x_t, \theta_0) \Rightarrow \theta = \theta_0.$$

Le pseudo-modèle est spécifié jusqu'au second ordre si, en plus de l'espérance, la variance conditionnelle de y_t sachant x_t , $V(y_t / x_t)$, permet de décrire la liaison entre y_t et x_t . La variance conditionnelle est alors connue et dépend des paramètres d'intérêt⁵ identifiables au premier ordre et d'autres paramètres (paramètres de nuisance) liés aux moments du second ordre, que nous noterons σ . Le pseudo modèle est spécifié comme suit :

$$\begin{aligned} y_t &= m(x_t, \theta) + u_t, \\ E(y_t / x_t, \theta) &\equiv m(x_t, \theta), \\ V(y_t / x_t, \theta) &= g(x_t, \theta, \sigma). \end{aligned}$$

(θ, σ) est identifiable au second ordre si pour toute valeur (θ_0, σ_0) :

$$\begin{aligned} m(x_t, \theta) = m(x_t, \theta_0) &\Rightarrow \theta = \theta_0 \\ g(x_t, \theta, \sigma) = g(x_t, \theta_0, \sigma_0) &\Rightarrow \sigma = \sigma_0 \end{aligned}$$

II- PRESENTATION GENERALE DU PMV

Soit le pseudo modèle défini à la section précédente :

$$y_t = m(x_t, \theta) + u_t,$$

où les termes d'erreur u_t , $t = 1, \dots, T$, sont distribués identiquement et indépendamment, et sont indépendants des variables exogènes x_t .

Si la vraie distribution de probabilité f_0 de u_t est connue, nous pouvons alors appliquer la méthode du maximum de vraisemblance pour estimer les paramètres θ . La log-vraisemblance, $\text{Ln}(L_0(\theta))$, à maximiser est alors de la forme :

$$\begin{aligned} \text{Ln}(L_0(\theta)) &= \frac{1}{T} \sum_{t=1}^T \text{Ln}(f_0(u_t)), \\ &= \frac{1}{T} \sum_{t=1}^T \text{Ln} [f_0(y_t - m(x_t, \theta))] \end{aligned}$$

où L_0 est la fonction de vraisemblance associée à la vraie distribution f_0 .

⁵Paramètres interprétables économiquement.

L'estimateur du maximum de vraisemblance $\hat{\theta}$ vérifie alors la condition de premier ordre suivante :

$$\frac{\partial \ln(L(\hat{\theta}))}{\partial \hat{\theta}} = 0.$$

On démontre, dans ce cas, que les estimateurs $\hat{\theta}$ sont convergents et asymptotiquement normaux.

Cependant, dans le cas général, la vraie distribution de probabilité est inconnue. La méthode du PMV consiste alors à maximiser la vraisemblance, pour une distribution de probabilité donnée f , différente de f_0 , établie à partir d'une connaissance partielle de cette vraie distribution f_0 . On s'intéresse donc aux estimateurs du maximum de vraisemblance calculés en faisant une hypothèse fictive sur la loi des perturbations, d'où le qualificatif *pseudo* (ou *quasi*).

L'étude des propriétés du PMV revient à l'analyse du maximum de vraisemblance en présence d'une erreur de spécification sur la loi des erreurs. La pseudo-log-vraisemblance est alors de la forme :

$$\begin{aligned} \ln(L(\theta)) &= \frac{1}{T} \sum_{i=1}^T \ln(f(u_i)), \\ &= \frac{1}{T} \sum_{i=1}^T \ln[f(y_i - m(x_i, \theta))] \end{aligned}$$

L'estimateur du PMV, $\hat{\theta}$, vérifie la condition :

$$\frac{\partial \ln(L(\hat{\theta}))}{\partial \hat{\theta}} = 0.$$

II.1- Information discriminante de Kullback et convergence du PMV : Analyse de White (1982)

En présence d'une erreur de spécification sur la loi de distribution des erreurs, il est intéressant de chercher la loi f la plus proche de f_0 , de manière à pouvoir évaluer cette erreur en comparant f_0 à f . Il faut pour cela disposer d'une mesure de proximité (appelée aussi contraste) entre lois de probabilités.

Etant données $f_0(x,y)$, la vraie loi des variables exogènes et des variables endogènes, et $f(x, y, \theta)$ la pseudo-densité du couple (y,x) ; on appelle contraste au sens de Kullback entre f et f_0 la quantité suivante :

$$I(f_0, f, \theta) = E_x E_0 \left[\ln \left(\frac{f_0(x, y)}{f(x, y, \theta)} \right) \right] \\ = \int \ln(f_0(x, y)) dF_0(x, y) - \int \ln(f(x, y, \theta)) dF_0(x, y)$$

où F_0 désigne la vraie fonction de distribution, E_0 représente l'espérance par rapport à la vraie distribution conditionnelle de y sachant x , et E_x correspond à l'espérance par rapport à la vraie loi marginale de x .

$I(\cdot)$ est aussi appelée information discriminante au sens de Kullback.

Sous des hypothèses générales (Annexe 1), White (1982) montre que les estimateurs du PMV convergent, presque sûrement, vers les paramètres qui minimisent l'information discriminante de Kullback. En d'autres termes, les estimateurs du PMV minimisent la "distance"⁶ entre le modèle $f(y-m(x, \theta))$ et la réalité $f_0(y-m(x, \theta))$.

Si $f(u, \theta) = f_0(u)$, alors le minimum de l'information de Kullback est atteint pour $\theta^* = \theta_0$. L'estimateur du PMV, $\hat{\theta}_T$, n'est autre que l'estimateur du maximum de vraisemblance.

L'estimateur du PMV est normalement distribué

$$\sqrt{T}(\tilde{\theta}_T - \theta^*) \xrightarrow{d} N(0, J_{-1}(\theta^*) I(\theta^*) J_{-1}(\theta^*))$$

De plus $C_T(\tilde{\theta}_T) \xrightarrow{p} C(\theta^*)$, élément par élément, avec $C(\theta^*) = J_{-1}(\theta^*) I(\theta^*) J_{-1}(\theta^*)$

$$J(\theta^*) = -E_x E_0 \left(\frac{\partial^2 \ln(f(u, \theta^*))}{\partial \theta \partial \theta} \right)$$

$$I(\theta^*) = -E_x E_0 \left(\frac{\partial \ln(f(u, \theta^*))}{\partial \theta} \frac{\partial \ln(f(u, \theta^*))}{\partial \theta} \right)$$

$$\text{et } C_T(\tilde{\theta}_T) = \tilde{J}_T^{-1} \tilde{I}_T \tilde{J}_T^{-1},$$

$$\tilde{J}_T = -\frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \ln(f(u_t, \tilde{\theta}_T))}{\partial \theta \partial \theta}$$

$$\tilde{I}_T = -\frac{1}{T} \sum_{t=1}^T \frac{\partial \ln(f(u_t, \tilde{\theta}_T))}{\partial \theta} \frac{\partial \ln(f(u_t, \tilde{\theta}_T))}{\partial \theta}$$

L'estimateur convergent de la matrice de variance-covariance des paramètres est la matrice $C_T(\hat{\theta}_T)$, où les espérances sont remplacées par les moyennes empiriques.

⁶Il ne s'agit pas d'une distance au sens classique du terme, puisque les conditions de symétrie et d'inégalité triangulaire ne sont pas satisfaites.

Rappelons que $\hat{\theta}_T$ est l'estimateur du PMV et θ^* le minimum unique de l'information de Kullback.

Toutefois, si les hypothèses sur la distribution suivie par les erreurs sont inadéquates ou si le modèle n'est pas correctement paramétré, alors θ^* peut ne pas coïncider avec la vraie valeur des paramètres. Ce qui signifie, qu'en règle générale, l'estimateur du PMV n'est pas convergent.

C'est principalement à ce niveau qu'interviennent les apports de Gouriéroux, Monfort et Trognon (1984). Les auteurs définissent en effet, les conditions sur la famille des pseudo-lois pour que les estimateurs du PMV ne perdent pas certaines de leurs propriétés comme la convergence ou l'efficacité asymptotique.

II.2- L'apport de Gouriéroux, Monfort et Trognon (1984)

Les auteurs montrent qu'une condition nécessaire et suffisante de convergence des estimateurs du PMV est que la loi des variables endogènes du pseudo-modèle appartienne à une famille exponentielle linéaire ou quadratique selon l'ordre d'identifiabilité du modèle.

A- Pseudo-Maximum de Vraisemblance d'ordre un (PMV1)

On se place d'abord dans le cas où le modèle est connu au premier ordre seulement. On retient donc l'écriture suivante avec une variance conditionnelle de y sachant x , $\Omega_0(x_t)$, inconnue :

$$\begin{aligned} y_t &= m(x_t, \theta_0) + u_t, \\ \text{avec } E(y_t / x_t) &\equiv m(x_t, \theta_0), \\ \text{et } V(y_t / x_t) &= \Omega_0(x_t) \end{aligned}$$

Sous les hypothèses H1 à H7 (annexe 1), il existe une condition nécessaire et suffisante sur la famille des lois conditionnelles de y_t sachant x_t qui assure la convergence de l'estimateur du PMV1 vers la vraie valeur, quelle que soit la loi sous-jacente :

L'estimateur $\hat{\theta}_T$ du PMV1, solution du problème :

$$\max_{\theta} \left\{ \frac{1}{T} \sum_{t=1}^T \ln f[y_t, m(x_t, \theta)] \right\},$$

est convergent si et seulement si la distribution retenue appartient à une famille exponentielle linéaire, c'est-à-dire une famille de lois dont la densité s'écrit sous la forme :

$$f(y_t, m(x_t, \theta)) = \exp[A(m(x_t, \theta)) + B(y_t) + C(m(x_t, \theta))y_t];$$

où $A(m(x_t, \theta))$ et $B(y_t)$ représentent des scalaires et $C(m(x_t, \theta))$ un vecteur ligne de dimension G .

On remarque que les contraintes associées aux variables endogènes ($B(y_t)$) sont indépendantes des paramètres (θ). Il en résulte que le critère à maximiser peut être réécrit :

$$\max_{\theta} \left\{ \frac{1}{T} \sum_{t=1}^T (A(m(x_t, \theta)) + C(m(x_t, \theta))y_t) \right\}$$

De nombreuses distributions, couramment utilisées en économétrie, appartiennent à la famille de lois exponentielles linéaires : les distributions de poisson, binomiale, binomiale négative, de gamma, multinomiale, normale.

Ainsi, pour une loi normale unidimensionnelle, on peut écrire :

$$f(y, m(x_t, \theta)) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{y - m(x_t, \theta)}{\sigma} \right)^2 \right\} \text{ pour } \sigma \text{ donné.}$$

Cette densité peut se réécrire :

$$f(y_t, m(x_t, \theta)) = \exp\left\{ -Ln(\sigma\sqrt{2\pi}) - \frac{1}{2} \frac{m(x_t, \theta)^2}{\sigma^2} - \frac{1}{2\sigma^2} y_t^2 + \frac{m(x_t, \theta)}{\sigma^2} y_t \right\},$$

et l'on posera :

$$A(m(x_t, \theta)) = -Ln(\sigma\sqrt{2\pi}) - \frac{1}{2} \frac{m(x_t, \theta)^2}{\sigma^2}; B(y_t) = -\frac{1}{2\sigma^2} y_t^2 \text{ et } C(m(x_t, \theta)) = \frac{m(x_t, \theta)}{\sigma^2}.$$

Un modèle simple à équations simultanées permet d'illustrer le principe du PMV1. Soit le système :

$$\begin{cases} y_{1,t} = \alpha_1 y_{2,t} + \beta_1 x_{1,t} + u_{1,t} \\ y_{2,t} = \alpha_2 y_{1,t} + \beta_2 x_{2,t} + u_{2,t} \end{cases}$$

avec $y_t = (y_{1,t}, y_{2,t})'$, $u_t = (u_{1,t}, u_{2,t})'$, $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2)'$, $x_t = (x_{1,t}, x_{2,t})$ et \mathfrak{J} le jacobien de la transformée de y_t en u_t : $\mathfrak{J} = \begin{pmatrix} 1 & -\alpha_1 \\ -\alpha_2 & 1 \end{pmatrix}$.

Sous l'hypothèse d'une matrice de variance-covariance Ω_0 connue et indépendante des paramètres θ , la log-vraisemblance du modèle prend la forme suivante:

$$-Ln(2\pi) - \frac{1}{2}Ln(\det(\Omega_0)) + Ln(\det(\mathfrak{I})) - \frac{1}{2}u_t'\Omega_0^{-1}u_t.$$

Si on note $H(y_t, x_t, \theta) = \begin{pmatrix} \alpha_1 y_{2,t} + \beta_1 x_{1,t} \\ \alpha_2 y_{1,t} + \beta_2 x_{2,t} \end{pmatrix}$ alors $u_t = y_t - H(y_t, x_t, \theta)$ et on obtient :

$$\begin{cases} A(H(y_t, x_t, \theta)) = -Ln(2\pi) - \frac{1}{2}Ln(\det(\Omega_0)) + Ln(\det(\mathfrak{I})) - \frac{1}{2}H'(y_t, x_t, \theta)\Omega_0^{-1}H(y_t, x_t, \theta) \\ B(y_t) = -\frac{1}{2}y_t'\Omega_0^{-1}y_t \\ C(H(y_t, x_t, \theta)) = H'(y_t, x_t, \theta)\Omega_0^{-1} \end{cases}$$

On peut réécrire ce système sous une forme équivalente :

$$\begin{cases} A(m(x_t, \theta)) = -Ln(2\pi) - \frac{1}{2}Ln(\det(\Sigma_0)) - \frac{1}{2}m'(x_t, \theta)\Sigma_0^{-1}m(x_t, \theta) \\ B(y_t) = -\frac{1}{2}y_t'\Sigma_0^{-1}y_t \\ C(m(x_t, \theta)) = m'(x_t, \theta)\Sigma_0^{-1} \end{cases}$$

où $m(x_t, \theta) = \begin{pmatrix} \frac{\alpha_1 \beta_2 x_{2,t} + \beta_1 x_{1,t}}{1 - \alpha_1 \alpha_2} \\ \frac{\alpha_2 \beta_1 x_{1,t} + \beta_2 x_{2,t}}{1 - \alpha_1 \alpha_2} \end{pmatrix}$ soit encore $E(y_t/x_t, \theta)$ et $\Sigma_0 = \mathfrak{I}^{-1}\Omega_0\mathfrak{I}^{-1}$, qui est la matrice

de variance-covariance des endogènes.

On retrouve le résultat selon lequel le PMV1 revient à maximiser le critère suivant :

$$A(m(x_t, \theta)) + C(m(x_t, \theta))y_t,$$

lequel critère ne fait plus intervenir le jacobien \mathfrak{I} de la transformée. Celui-ci est directement incorporé dans l'évaluation des moments de la variable endogène.

Bien entendu cet exemple n'a qu'un but illustratif : le transfert du jacobien de la vraisemblance vers les moments de la variable endogène n'apporte pas de simplifications notables. En revanche, pour certains modèles fortement non-linéaires, une telle démarche peut être cruciale pour obtenir une expression manipulable de la vraisemblance.

Remarque : *L'estimateur des moindres carrés non-linéaires correspond à un cas particulier du PMV1.*

Encadré 1 : Normalité asymptotique de l'estimateur du PMV1

L'estimateur du PMV1 associé à la famille exponentielle linéaire est asymptotiquement normal :

$$\sqrt{T}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} N(0, J^{-1}IJ^{-1})$$

avec :

$$J = E_x E_0 \left(-\frac{\partial^2 \ln[f(y, m(x, \theta_0))]}{\partial \theta \partial \theta} \right),$$

$$J = E_x \left(\frac{\partial m(x, \theta_0)' \partial C[m(x, \theta_0)] \partial m(x, \theta_0)}{\partial \theta \partial m(\theta_0) \partial \theta} \right),$$

$$I = E_x E_0 \left(\frac{\partial \ln[f(y, m(x, \theta_0))]}{\partial \theta} \frac{\partial \ln[f(y, m(x, \theta_0))]}{\partial \theta} \right),$$

$$I = E_x \left(\frac{\partial m(x, \theta_0)' \partial C[m(x, \theta_0)]}{\partial \theta \partial m(\theta_0)} E_0 \{ u(\theta_0) u(\theta_0)' \} \frac{\partial C[m(x, \theta_0)] \partial m(x, \theta_0)}{\partial m(\theta_0) \partial \theta} \right)$$

avec :

$$u(\theta) = y - m(x, \theta)$$

L'utilisation de la famille exponentielle linéaire simplifie sensiblement l'expression des matrices J (matrice d'information de Fisher) et I (produit du vecteur du score). Ainsi, on peut exprimer J et I comme suit :

$$J = E_x \left(\frac{\partial m(x, \theta_0)' \Sigma_0^{-1} \partial m(x, \theta_0)}{\partial \theta \partial \theta} \right),$$

$$I = E_x \left(\frac{\partial m(x, \theta_0)' \Sigma_0^{-1} \Omega_0 \Sigma_0^{-1} \partial m(x, \theta_0)}{\partial \theta \partial \theta} \right),$$

avec $\Omega_0 = E_0(u(\theta_0)u(\theta_0)')$;

Ω_0 est la matrice de variance-covariance conditionnelle de y sachant x pour la vraie distribution.

Les estimateurs convergents de J et I sont respectivement :

$$\tilde{J}_T = \frac{1}{T} \sum_{t=1}^T \left(\frac{\partial m(x, \tilde{\theta}_t)' \partial \mathcal{C}[m(x, \tilde{\theta}_t)]}{\partial \theta} \frac{\partial m(x, \tilde{\theta}_t)}{\partial \theta} \right)$$

$$\tilde{I}_T = \frac{1}{T} \sum_{t=1}^T \left(\frac{\partial m(x, \tilde{\theta}_t)' \partial \mathcal{C}[m(x, \tilde{\theta}_t)]}{\partial \theta} \left\{ \mu_t(\tilde{\theta}_t) \mu_t(\tilde{\theta}_t)' \right\} \frac{\partial \mathcal{C}[m(x, \tilde{\theta}_t)]}{\partial m(\tilde{\theta}_t)} \frac{\partial m(x, \tilde{\theta}_t)}{\partial \theta} \right)$$

B- Le PMV Quasi Généralisé (PMVQG)

Pour un modèle identifié jusqu'au second ordre, on peut tirer profit de l'information supplémentaire contenue dans la variance conditionnelle, $V(y_t/x_t) = g(x_t, \theta, \sigma)$ (supposée connue de l'expérimentateur), pour estimer de manière plus précise le paramètre d'intérêt θ . Dans ce cas, les familles exponentielles linéaires sont paramétrées à la fois par le vecteur des coefficients liés au moment du premier ordre (θ) et par celui lié aux paramètres identifiables au second ordre (les paramètres de nuisance μ). Le vecteur μ est fonction des moments du premier et du second ordre :

$$\mu = h(m(x, \theta), g(x, \theta, \sigma)).$$

Pour $m(x, \theta)$ donné, la relation entre μ et $g(x, \theta, \sigma)$ est bijective.

La méthode du PMVQG procède en deux étapes :

- Obtenir, dans un premier temps, des estimateurs convergents $\hat{\sigma}$ et $\hat{\theta}$ de σ et θ qui permettent de construire un estimateur convergent du paramètre de nuisance $\mu = h(m(x, \hat{\theta}), g(x, \hat{\theta}, \hat{\sigma}))$;
- construire la matrice de variances-covariances $V(y_t/x_t)$ et estimer les paramètres identifiables au premier ordre θ .

Remarque : L'estimateur du PMVQG résulte de l'adaptation de la technique des moindres carrés quasi-généralisés en deux étapes au cas du pseudo-maximum de vraisemblance (Trognon (1987)).

Soient $\hat{\sigma}$ et $\hat{\theta}$ deux estimateurs fortement convergents de σ_0 et θ_0 et tels que $\sqrt{T}(\hat{\theta}_T - \theta_0)$ et $\sqrt{T}(\hat{\sigma}_T - \sigma_0)$ sont bornés en probabilité. L'estimateur $\hat{\theta}_T$ de θ du PMVQG, obtenu en maximisant le critère :

$$\sum_{t=1}^T \text{Ln} f \left[y_t, m(x_t, \theta), h \left(m(x_t, \tilde{\theta}_T), g(x_t, \tilde{\theta}_T, \tilde{\sigma}_T) \right) \right]$$

est fortement convergent et est le meilleur estimateur asymptotiquement normal du PMV pour une famille exponentielle linéaire :

$$\sqrt{T}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} N(0, \kappa_0^{-1}) \text{ où } \kappa_0 = E_x \left(\frac{\partial m(x, \theta_0)}{\partial \theta} \Omega_0^{-1} \frac{\partial m(x, \theta_0)}{\partial \theta} \right).$$

Gouriéroux, Monfort et Trognon montrent que l'estimateur du PMVQG est asymptotiquement plus efficace que l'estimateur du PMV1.

Les auteurs montrent par ailleurs, que si θ et σ sont fonctionnellement indépendants et sous des hypothèses de régularité, alors l'estimateur $\hat{\theta}_T$ du PMVQG de θ_0 est asymptotiquement équivalent à l'estimateur du MV, solution du problème :

$$\max_{\theta, \sigma} \left\{ \frac{1}{T} \sum_{t=1}^T \text{Ln} f_0(y_t, x_t, \theta, \sigma) \right\}.$$

C- Le PMV d'ordre deux (PMV2)

Pour un modèle identifiable au second ordre, les paramètres d'intérêt (θ) et de nuisance (σ) peuvent être estimés simultanément par la méthode du PMV2.

Sous les hypothèses H1 à H7 (annexe 1), l'estimateur $(\hat{\theta}_T, \hat{\sigma}_T)$ du PMV2, solution du problème :

$$\max_{\theta, \sigma} \left\{ \frac{1}{T} \sum_{t=1}^T \text{Ln} f((y_t, m(x_t, \theta), g(x_t, \sigma))) \right\},$$

est un estimateur fortement convergent de la vraie valeur des paramètres (θ_0, σ_0) si la famille de loi conditionnelle appartient à une famille exponentielle quadratique, c'est-à-dire une famille dont la densité s'écrit sous la forme :

$$f(y, m(x, \theta), g(x, \sigma)) = \exp \left\{ A(m(x, \theta), g(x, \sigma)) + B(y) + C(m(x, \theta), g(x, \sigma))y + y' D(m(x, \theta), g(x, \sigma))y \right\}$$

où $D(m(x, \theta), g(x, \sigma))$ est une matrice carrée de dimension $G \times G$.

Comme pour l'estimateur du PMV1, l'estimateur du PMV2 est asymptotiquement normal :

$$\sqrt{T} \begin{bmatrix} \tilde{\theta}_T \\ \tilde{\sigma}_T \end{bmatrix} - \begin{bmatrix} \theta_0 \\ \sigma_0 \end{bmatrix} \xrightarrow{d} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, J^{-1} J J^{-1} \right)$$

Cependant, les expressions des matrices I et J sont sensiblement plus compliquées (voir annexe 1).

Telles que l'on vient de les présenter, les méthodes du PMV reposent sur la connaissance exacte de la forme fonctionnelle du premier moment conditionnel (pour PMV1 et PMVQG) ou des deux premiers moments conditionnels de la variable endogène (pour PMV2). Si ces méthodes simplifient notablement l'expression de la vraisemblance à maximiser, elles n'affranchissent pas l'économètre du calcul de l'expression des deux premiers moments. Or pour certains modèles économétriques, le calcul effectif de ces moments est très difficile, voire impossible. Laroque et Salanié (1989) proposent alors d'utiliser des simulations stochastiques pour approximer empiriquement les espérances et les variances non calculables. Ils présentent ainsi une version simulée du PMV.

III- PRESENTATION DE LA VERSION SIMULEE DU PMV

III.1- Bref rappel historique des méthodes simulées

Traditionnellement, les méthodes de simulation sont largement utilisées pour évaluer empiriquement les propriétés d'une méthode d'estimation, ou d'une procédure de test : étudier le biais et l'erreur quadratique moyenne d'un estimateur ou le niveau et la puissance d'un test statistique, calculer des prévisions sans biais dans le cas des modèles non linéaires, ou approximer des résidus latents dans les modèles à variables qualitatives.

Au cours des dix dernières années, plusieurs auteurs ont proposé une méthodologie d'estimation par simulation. Cette méthodologie permet d'approximer des intégrales multiples apparaissant dans des fonctions objectifs (fonctions de vraisemblance) ou de passer outre à l'absence de forme analytique de la fonction de vraisemblance. Lerman et Manski (1981) ont été les premiers à suggérer de remplacer des probabilités analytiques nécessaires au calcul de la fonction de vraisemblance du modèle probit multinomial par des fréquences empiriques simulées. Le développement remarquable des moyens de calcul informatique durant les années quatre-vingts a permis de fructifier et de généraliser les suggestions de Lerman et Manski. L'article de Pakes (1986) constitue la première application dans ce domaine.

L'utilisation et la généralisation des méthodes d'estimation par simulation a depuis évolué rapidement⁷ (McFadden (1989), Pakes et Pollard (1989), Hajivassiliou et McFadden (1989), Duffie et Singleton (1989), Laroque et Salanié

⁷Voir le numéro spécial du Journal of Applied Econometrics, Décembre 1993.

(1989,1990,1992), Gouriéroux et Monfort (1991), Danielson et Richard (1992), etc.).

III.2- Le Pseudo-Maximum de Vraisemblance Simulé (PMVS)

Les méthodes du PMV reposent sur l'information contenue dans les moments du premier et du second ordre de la variable endogène. Ces moments peuvent être calculés de manière analytique; ils peuvent aussi être calculés numériquement par une *méthode de simulation* ou *méthode de Monte-carlo*.

Un modèle économétrique non linéaire peut en toute généralité s'écrire sous sa forme réduite suivante :

$$y_t = \phi(x_t, \theta, u_t).$$

En pratique, les simulations stochastiques sont réalisées en remplaçant le vecteur des erreurs u_t par un vecteur $u_{t,h}$, $h = 1, \dots, H$, de variables aléatoires tirées dans la loi des u_t d'espérance nulle et de variance égale à Ω . L'indice h représente le h ème tirage parmi H .

Pour chaque tirage h , on peut donc résoudre le modèle $\phi(x_t, \theta, u_{t,h})$ pour des valeurs données de x_t et θ .

A partir de $\phi(x_t, \theta, u_{t,h})$ on peut calculer des estimateurs de l'espérance et de la variance conditionnelle de y sachant x_t et θ ⁸:

$$m_H(x_t, \theta) = \frac{1}{H} \sum_{h=1}^H \phi(x_t, \theta, u_{h,t})$$

$$\approx m(x_t, \theta) = E_u \phi(x_t, \theta, u)$$

et

$$g_H(x_t, \theta, \sigma) = \frac{1}{H-1} \sum_{h=1}^H (\phi(x_t, \theta, u_{h,t}) - m_H(x_t, \theta))^2$$

$$\approx g(x_t, \theta, \sigma) = V_u \phi(x_t, \theta, \sigma).$$

Ces estimateurs ne sont autres que la moyenne empirique (pour l'espérance) et la variance empirique (pour la variance).

L'estimateur du PMVS (du premier et du second ordre), $\hat{\theta}_{TH}$, résulte alors de la maximisation (par rapport à θ pour une méthode d'ordre un, et par rapport à θ et σ pour une méthode d'ordre deux) de la pseudo-log-vraisemblance simulée suivante :

⁸Il résulte de la loi des grands nombres que, si H est grand, la moyenne et la variance empiriques sont peu différentes respectivement de l'espérance et de la variance de la loi ayant servi au tirage.

- pour une méthode d'ordre un (PMVS1) :

$$\text{Ln}(L(\theta))_T^H = \frac{1}{T} \sum_{t=1}^T \text{Ln}(f(y_t, m^H(x_t, \theta))),$$

- pour une méthode d'ordre deux (PMVS2) :

$$\text{Ln}(L(\theta, \sigma))_T^H = \frac{1}{T} \sum_{t=1}^T \text{Ln}(f(y_t, m^H(x_t, \theta), g^H(x_t, \theta, \sigma))).$$

Laroque et Salanié (1989) montrent que les estimateurs du PMVS sont asymptotiquement convergents lorsque le nombre de simulations H tend vers l'infini quand T tend vers l'infini. Les auteurs prouvent également que sous des hypothèses de différentiabilité, $\hat{\theta}_{TH}$ est asymptotiquement normal et de même variance-covariance asymptotique que l'estimateur du PMV $\hat{\theta}_T$, si le nombre de simulations H tend vers l'infini plus vite que \sqrt{T} .

IV- APPLICATION AU CAS DES MODELES DE DESEQUILIBRE

IV.1- Présentation du modèle et des approches alternatives d'estimation

Les modèles de déséquilibre reviennent sur l'hypothèse de flexibilité parfaite des prix et proposent une approche alternative fondée sur l'existence d'un ajustement par les quantités. Cet ajustement peut être total, auquel cas l'équilibre se réalise à un prix donné au minimum de l'offre (y_s^*) et de la demande (y_d^*). y_s^* et y_d^* ne sont pas toujours observables (l'offre n'est observée qu'en période d'excès de demande et la demande n'est observée qu'en période d'excès d'offre), ce sont des variables latentes. Ainsi, le *modèle latent* est défini par :

$$(1) \quad \begin{cases} y_{st}^* = x_{st}\theta_s + \sigma_s u_{st}, \\ y_{dt}^* = x_{dt}\theta_d + \sigma_d u_{dt}, \end{cases}$$

où x_s et x_d représentent respectivement les variables exogènes déterminant l'offre et la demande; u_s et u_d désignent les erreurs définies sur l'offre et la demande respectivement; nous supposons dans la suite que ces erreurs suivent des lois normales centrées, réduites et indépendantes; $\theta' = (\theta_s, \theta_d)$ est le vecteur des paramètres d'intérêt et $\sigma' = (\sigma_s, \sigma_d)$, le vecteur des paramètres de nuisance. Le *modèle observable* est défini par l'équation de la transaction réalisée (y_t) :

$$(2) \quad y_t = \min(y_{st}^*, y_{dt}^*), t=1, \dots, T.$$

le modèle défini par les équations (1) et (2) est appelé *modèle canonique de déséquilibre* (Quandt (1982)).

Depuis les travaux fondateurs de Fair et Jaffe (1972), les méthodes d'estimation des modèles de déséquilibre ont connu un véritable essor. La contrainte majeure réside alors dans l'imperfection de l'information disponible sur le découpage de l'échantillon en période d'excès d'offre et en période d'excès de demande.

Les travaux de Maddala et Nelson (1974) vont permettre d'envisager l'estimation de tels modèles à l'aide de la méthode du maximum de vraisemblance, qui ne suppose pas une connaissance préalable du découpage de l'échantillon, mais qui soulève malgré tout d'autres difficultés (voir annexe 2).

IV.2- Estimation par le PMV et par le PMVS

Nous présenterons successivement l'estimation par le PMV et le PMVS, du modèle canonique de déséquilibre (1er cas), du modèle avec micro marchés (2ème cas) et du modèle dynamique de déséquilibre (3ème cas). Dans le premier cas, les expressions analytiques des deux premiers moments conditionnels de la variable endogène observée sont faciles à calculer. On peut donc appliquer les procédures du PMV d'ordre un et deux pour estimer les paramètres d'intérêt et de nuisance. Dans le deuxième et troisième cas, le calcul de l'espérance et de la variance conditionnelles de la variable endogène observée n'est pas aisé, et pose des problèmes numériques liés à l'évaluation d'intégrales multiples (intégrale double dans le cas de micro marchés, une intégrale d'ordre T dans le cas du modèle dynamique). On préfère alors utiliser les méthodes du PMVS qui fournissent, à un coût faible, des estimateurs qui ont de bonnes propriétés asymptotiques et qui sont robustes vis-à-vis de l'introduction des micro marchés, et d'une structure dynamique de l'offre et de la demande.

A- Modèle canonique de déséquilibre : Estimation par le PMV

Pour une date t, $t = 1, \dots, T$, le modèle s'écrit :

$$(1) \quad y_t = \min(x_{st}\theta_s + \sigma_s u_{st}; x_{dt}\theta_d + \sigma_d u_{dt})$$

Nous noterons dans la suite $\sigma = (\sigma_d, \sigma_s)$, $s = \sqrt{\sigma_d^2 + \sigma_s^2}$, $x = (x_s, x_d)$ et $\theta = (\theta_s, \theta_d)$. Nous supposerons que les aléas u_d et u_s sont normalement distribués d'espérances nulles et de variances unitaires.

Nous avons vu dans les sections précédentes, que les méthodes du PMV reposent sur les deux premiers moments de la variable endogène. La première étape consiste donc à calculer les expressions de ceux-ci. Pour cela, on utilise les résultats

relatifs aux moments d'une loi normale bi-dimensionnelle tronquée (Gouriéroux (1989)) :

1ère étape : les expressions des deux premiers moments de la variable y

$$(2) \quad E(y_t / x_t) \equiv m(x_t, \theta, s) = x_{dt}\theta_d \Phi\left(\frac{x_{st}\theta_s - x_{dt}\theta_d}{s}\right) + x_{st}\theta_s \Phi\left(\frac{x_{dt}\theta_d - x_{st}\theta_s}{s}\right) - s \varphi\left(\frac{x_{dt}\theta_d - x_{st}\theta_s}{s}\right)$$

$$(3) \quad E(y_t^2 / x_t) \equiv g(x_t, \theta, \sigma) = ((x_{dt}\theta_d)_2 + \sigma_d^2) \Phi\left(\frac{x_{st}\theta_s - x_{dt}\theta_d}{s}\right) + ((x_{st}\theta_s)_2 + \sigma_s^2) \Phi\left(\frac{x_{dt}\theta_d - x_{st}\theta_s}{s}\right) - s(x_{dt}\theta_d + x_{st}\theta_s) \varphi\left(\frac{x_{dt}\theta_d - x_{st}\theta_s}{s}\right)$$

où Φ et φ sont respectivement la fonction de répartition et la densité de la loi normale centrée et réduite⁹.

Les équations (2) et (3) montrent que seuls θ et s sont identifiables au premier ordre, tandis que θ et σ sont identifiables au second ordre.

2ème étape : Calcul de la pseudo-vraisemblance à la date t

• Méthode d'ordre un : PMV1

Cette méthode consiste à maximiser la fonction de vraisemblance associée au pseudo-modèle suivant :

$$y_t = m(x_t, \theta, s) + u_{1t},$$

où les u_{1t} sont des variables aléatoires indépendantes, d'espérances nulles et dont la densité appartient à une famille exponentielle quadratique, plus précisément dans notre cas la loi normale.

En faisant *comme si* la variable endogène (y) suivait une loi normale, de variance V fixe et indépendante des paramètres, pour une observation t donnée et pour un choix donné des paramètres, la pseudo-vraisemblance s'écrit :

$$l_t(y_t, x_t, \theta, s) = \varphi(y_t - m(x_t, \theta, s)) = \frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{(y_t - m(x_t, \theta, s))^2}{2V}\right)$$

⁹Rappelons qu'une variable aléatoire normale, centrée et réduite (x) a pour fonction de densité de probabilité la fonction $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ et pour fonction de répartition $\Phi(x) = \int_{-\infty}^x \varphi(z) dz$.

Où encore le logarithme de la pseudo-vraisemblance à la date t :

$$(4) \quad \text{Ln}(l_{1t}) = -\frac{(y_t - m(x_t, \theta, s))^2}{2V} - \frac{\text{Ln}(2\pi V)}{2}.$$

• Le PMVQG :

Si on dispose, pour une valeur arbitraire des variables exogènes, d'un estimateur convergent de la variance de la variable endogène, $V(x_t, \hat{\theta}, \hat{\sigma})$. Si (y_t) était normalement distribué, le logarithme de sa pseudo-vraisemblance pour un choix donné des paramètres est :

$$(5) \quad \text{Ln}(l_{QGt}) = -\frac{(y_t - m(x_t, \theta, s))^2}{2V(x_t, \tilde{\theta}, \tilde{\sigma})} - \frac{\text{Ln}(2\pi V(x_t, \tilde{\theta}, \tilde{\sigma}))}{2}.$$

• Méthode d'ordre deux : PMV2

Il s'agit de maximiser la fonction de vraisemblance associée au pseudo-modèle suivant :

$$y_t = m(x_t, \theta, s) + u_{2t},$$

où les u_{2t} sont des perturbations normales, indépendantes, centrées et de variance $V(x_t, \theta, \sigma)$. Pour une observation donnée et pour un choix des paramètres, le logarithme de la pseudo-vraisemblance est :

$$(6) \quad \text{Ln}(l_{2t}) = -\frac{(y_t - m(x_t, \theta, s))^2}{2V(x_t, \theta, \sigma)} - \frac{\text{Ln}(2\pi V(x_t, \theta, \sigma))}{2}.$$

3ème étape : Calculer la pseudo-log-vraisemblance totale

Sous l'hypothèse d'indépendance des observations, la pseudo-log-vraisemblance totale correspond à la somme des pseudo-log-vraisemblances élémentaires, calculées à la deuxième étape :

• L'estimateur du PMV1 est solution de la maximisation de la fonction :

$$(7) \quad L_1 = \sum_{t=1}^T \text{Ln}(l_{1t}),$$

ou encore de la minimisation du critère suivant :

$$\sum_{t=1}^T (y_t - m(x_t, \theta, s))^2$$

• l'estimateur du PMVQG est solution de la maximisation de la fonction :

$$(8) \quad L_{QG} = \sum_{t=1}^T \text{Ln}(l_{QGt}),$$

ou encore de la minimisation du critère suivant :

$$\sum_{t=1}^T \frac{(y_t - m(x_t, \theta, s))^2}{V(x_t, \tilde{\theta}, \tilde{\sigma})}$$

• l'estimateur du PMV2 est solution de la maximisation de la fonction :

$$(9) \quad L_2 = \sum_{t=1}^T \text{Ln}(l_2t).$$

ou encore de la minimisation du critère suivant :

$$\sum_{t=1}^T \left(\frac{(y_t - m(x_t, \theta, s))^2}{V(x_t, \theta, \sigma)} + \text{Ln}(V(x_t, \theta, \sigma)) \right)$$

4ème étape : Détermination de l'optimum

Les paramètres d'intérêt et/ou de nuisances estimés sont issus de la minimisation de l'une des fonctions objectifs, définies à l'étape précédente. Cette minimisation est effectuée par un algorithme standard d'optimisation (méthode de Newton-Raphson, méthode de la montée la plus rapide, méthode du scoring ...etc.).

B- Modèles de déséquilibre avec micro marchés : Estimation par le PMVS

Si l'agrégation de micro marchés (voir annexe 2.B) permet de tenir compte de la concomitance des régimes d'excès d'offre et d'excès de demande, elle complique néanmoins le calcul des premiers moments conditionnels de la variable endogène (la transaction agrégée). Dans cette situation, la méthode du PMV est d'application aussi délicate que celle du MV. Pour résoudre ce problème, on utilise la version simulée du PMV.

On fait l'hypothèse que les micro marchés indicés par i , $i = 1, \dots, N$, sont infiniment petits et infiniment nombreux (un continuum de micro marchés). Sous cette hypothèse, l'échange au niveau agrégé est égale à l'espérance des transactions sur les micro marchés :

$$(10) \quad y_t = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \min(x_{st}\theta_s + \sigma_s u_{st} + \tau_s v_{si}; x_{dt}\theta_d + \sigma_d u_{dt} + \tau_d v_{di}) \\ = E_i(\min(x_{st}\theta_s + \sigma_s u_{st} + \tau_s v_{si}; x_{dt}\theta_d + \sigma_d u_{dt} + \tau_d v_{di}) / u_{st}, u_{dt})$$

où l'espérance mathématique (E_i) est prise par rapport à (v_{si}, v_{di}) .

On considère donc qu'il y a des perturbations qui affectent l'ensemble des micro marchés à la date t (il s'agit de u_{st} et u_{dt}), alors que la particularité du micro-marché i est prise en compte par les termes d'erreur v_{si} et v_{di} . On suppose dans la suite que ces perturbations sont normalement distribuées, d'espérances nulles, de variances unitaires, non autocorrélées (dans le temps pour u_{st} et u_{dt} , et dans l'espace pour v_{si} et v_{di}) et non corrélées entre elles.

Sous les hypothèses mentionnées ci-dessus, l'espérance mathématique de la transaction réalisée (équation (10)) s'écrit :

$$(11) \quad E_i(y_t) = (x_{dt}\theta_d + \sigma_d u_{dt}) \Phi\left(\frac{x_{st}\theta_s + \sigma_s u_{st} - x_{dt}\theta_d - \sigma_d u_{dt}}{\tau}\right) \\ + (x_{st}\theta_s + \sigma_s u_{st}) \Phi\left(\frac{x_{dt}\theta_d + \sigma_d u_{dt} - x_{st}\theta_s - \sigma_s u_{st}}{\tau}\right) - \tau \varphi\left(\frac{x_{dt}\theta_d + \sigma_d u_{dt} - x_{st}\theta_s - \sigma_s u_{st}}{\tau}\right)$$

où Φ et φ désignent respectivement la fonction de répartition et la fonction de densité de la loi normale centrée et réduite. On notera que le dernier terme à droite de l'équation (11), peut être considéré comme un indicateur des frictions observées sur les micro marchés; $\tau = \sqrt{\tau_d^2 + \tau_s^2}$ représente l'indicateur de dispersion de l'écart entre l'offre et la demande sur les micro marchés.

On déduit de l'équation (11) que :

- les paramètres $(\theta_s, \theta_d, s_2 + \tau_2)$ sont identifiables au premier ordre. Le calcul de la variance de y_t est plus laborieux, il montre que les paramètres $(\theta_d, \theta_s, \sigma_d, \sigma_s, \tau)$ sont identifiables au second ordre;

- Puisque τ et s ne sont pas identifiables séparément, utiliser le PMV1 en l'absence ou en présence de micro-marché conduit aux mêmes résultats et n'est donc d'aucune utilité pour estimer le modèle avec micro marchés. Seul le PMV2 présente ici un intérêt.

Pour obtenir l'espérance de la quantité échangée au niveau agrégé, on doit intégrer l'équation (11) par rapport à la densité de (u_{dt}, u_{st}) . D'où l'intérêt d'approximer les deux premiers moments par simulation.

1ère étape : Approximation des deux premiers moments conditionnels en procédant à des simulations de Monte carlo

Nous procédons, à chaque période, au tirage de H (entre 50 et 100 dans la pratique) perturbations normales centrées, indépendantes et de variances unitaires que nous noterons :

$$(u_{h,t}^s, u_{h,t}^d),$$

Nous conservons les mêmes séries tout au long de la procédure d'estimation afin d'opérer avec la même approximation des fonctions de la vraisemblance (selon l'ordre d'identifiabilité du modèle).

Pour $(\theta_d, \theta_s, \sigma_d, \sigma_s, \tau)$ donnés (ce qui revient à fixer ses paramètres), nous calculons :

$$y_{h,t} = (x_{dt}\theta_d + \sigma_d u_{h,t}^d) \Phi\left(\frac{x_{st}\theta_s + \sigma_s u_{h,t}^s - x_{dt}\theta_d - \sigma_d u_{h,t}^d}{\tau}\right) \\ + (x_{st}\theta_s + \sigma_s u_{h,t}^s) \Phi\left(\frac{x_{dt}\theta_d + \sigma_d u_{h,t}^d - x_{st}\theta_s - \sigma_s u_{h,t}^s}{\tau}\right) - \tau \phi\left(\frac{x_{dt}\theta_d + \sigma_d u_{h,t}^d - x_{st}\theta_s - \sigma_s u_{h,t}^s}{\tau}\right) \\ R_{t,h} = \Phi\left(\frac{x_{st}\theta_s + \sigma_s u_{h,t}^s - x_{dt}\theta_d - \sigma_d u_{h,t}^d}{\tau}\right)$$

puis nous approximations les moments de la variable endogène par :

$$E(y_t/x_t) = \frac{\sum_{h=1}^H y_{h,t}}{H} \quad \text{et} \quad E(y_t^2/x_t) = \frac{\sum_{h=1}^H y_{h,t}^2}{H-1}.$$

Les proportions de micro marchés en régimes d'offre et de demande sont approximées par :

$$\text{Pr}(\text{régime de demande}) = \frac{\sum_{h=1}^H R_{h,t}}{H} \quad \text{et} \quad \text{Pr}(\text{régime d'offre}) = 1 - \frac{\sum_{h=1}^H R_{h,t}}{H}.$$

Connaissant les deux premiers moments, la mise en oeuvre de la méthode du PMVS d'ordre 2 est identique à la description fournie dans le cas du modèle canonique de déséquilibre¹⁰ :

2ème étape : *Calcul de la vraisemblance à la date t,*

3ème étape : *Calculer la log-vraisemblance totale,*

4ème étape : *Détermination de l'optimum.*

C- Modèles de déséquilibre dynamiques : Estimation par le PMVS

L'un des griefs adressé aux modèles de déséquilibre est la nature généralement sommaire de la spécification dynamique retenue. La présence de variables latentes retardées (soit d'une dynamique sur les variables non observables, l'offre et la demande) trouve pourtant de multiples justifications .

Cette carence en matière de dynamique a pour origine essentiellement des difficultés numériques d'estimation liées à la complexité de l'expression analytique de la fonction de vraisemblance. Dans ce contexte, les méthodes du PMVS se sont avérées particulièrement fécondes puisqu'elles se prêtent aisément à une extension dynamique (Laroque et Salanié (1992, 1993), Ben Jelili et Mihoubi (1993)). Les méthodes du PMV Dynamique Simulé (PMVDS) sont fondées sur l'évaluation des différents moments de la variable endogène observée par des simulations dynamiques : les variables latentes retardées prennent pour valeurs à la période t celles qui ont été simulées par les variables latentes aux périodes précédentes.

Présentation du modèle

Nous considérons le modèle stationnaire¹¹ suivant :

$$\begin{aligned} y_{st}^* &= \rho_s y_{st-1}^* + x_{st} \theta_s + \sigma_s u_{st}, \\ y_{dt}^* &= \rho_d y_{dt-1}^* + x_{dt} \theta_d + \sigma_d u_{dt}, \\ y_t &= \min(y_{st}^*, y_{dt}^*). \end{aligned}$$

¹⁰La maximisation de la pseudo-log-vraisemblance requiert la connaissance des dérivées des deux premiers moments par rapport aux paramètres. Les approximations de ces dérivées par la méthode de Monte-Carlo sont présentées dans l'article de Laffargue et Mihoubi (1989).

¹¹Dans les applications concrètes, les variables macro-économiques sont généralement non stationnaires. Les conséquences de l'absence de stationnarité quant à l'applicabilité du PMVS sont analysées dans Laroque et Salanié (1992) et Mihoubi (1993).

avec $|\rho_s| < 1$ et $|\rho_d| < 1$.

Les perturbations (u_{st}, u_{dt}) sont supposées normalement et indépendamment distribuées, et non corrélées entre elles.

Nous pouvons réécrire le modèle, par récurrence, sous sa forme finale :

$$\begin{aligned} y_{st}^* &= \rho_s^t y_{s0}^* + \theta_s \sum_{j=0}^{t-1} \rho_s^j x_{st-j} + \sigma_s \sum_{j=0}^{t-1} \rho_s^j u_{st-j}, \\ y_{dt}^* &= \rho_d^t y_{d0}^* + \theta_d \sum_{j=0}^{t-1} \rho_d^j x_{dt-j} + \sigma_d \sum_{j=0}^{t-1} \rho_d^j u_{dt-j}, \\ y_t &= \min(y_{st}^*, y_{dt}^*). \end{aligned}$$

y_{s0}^* et y_{d0}^* représentent les valeurs initiales respectivement de l'offre et de la demande.

Sous cette forme, les paramètres liés au moment du premier ordre $(\theta_s, \theta_d, \rho_s, \rho_d)$ sont identifiables avec les trois méthodes du PMV. Les paramètres (σ_s, σ_d) ne sont identifiables qu'avec le PMV2, puisque seule cette méthode d'estimation tient compte des moments de second ordre.

Simulation des deux premiers moments conditionnels

On effectue, à chaque période t , H tirages de perturbations auxiliaires normales, centrées, réduites, indépendantes et non autocorrélées : $u_{h,t}^s, u_{h,t}^d$. Ces dernières doivent être conservées tout au long de la procédure d'optimisation de la vraisemblance. Pour simplifier la présentation, nous supposons que les valeurs initiales des variables latentes (y_{s0}^*, y_{d0}^*) sont connues.

Pour un tirage h , $h = 1, \dots, H$, nous pouvons calculer $(y_{sh,t}^*, y_{dh,t}^*)$ par simulation dynamique :

$$\begin{aligned} y_{sh,t}^* &= \rho_s y_{sh,t-1}^* + x_{st} \theta_s + \sigma_s u_{h,t}^s, \\ y_{dh,t}^* &= \rho_d y_{dh,t-1}^* + x_{dt} \theta_d + \sigma_d u_{h,t}^d, \end{aligned}$$

avec :

$$\begin{aligned} y_{sh,0}^* &= y_{s0}, \\ y_{dh,0}^* &= y_{d0}. \end{aligned}$$

De la même façon, nous définissons :

$$R_{h,t} = \begin{cases} 1 & \text{si } y_{dh,t}^* > y_{sh,t}^* \\ 0 & \text{sinon} \end{cases}$$

$$y_{h,t} = y_{dh,t}^*(1 - R_{h,t}) + y_{sh,t}^* R_{h,t}.$$

Ainsi, nous obtenons les approximations des deux premiers moments¹² :

$$E(y_t / \underline{x}_t, y_{s0}^*, y_{d0}^*, \underline{u}_{st}, \underline{u}_{dt}) = \frac{1}{H} \sum_{h=1}^H y_{h,t},$$

$$E(y_t^2 / \underline{x}_t, y_{s0}^*, y_{d0}^*, \underline{u}_{st}, \underline{u}_{dt}) = \frac{1}{H-1} \sum_{h=1}^H y_{h,t}^2.$$

où $\underline{x}_t = (x_1, \dots, x_t)$, $\underline{u}_{st} = (u_{s1}, \dots, u_{st})$ et $\underline{u}_{dt} = (u_{d1}, \dots, u_{dt})$.

Ces approximations serviront à entamer la procédure d'estimation par le PMV telle qu'elle est décrite dans le cas canonique (les étapes 2,3 et 4).

IV.3- Application à l'emploi en déséquilibre et en présence de variables latentes retardées.

L'exemple illustratif que nous allons présenter concerne le marché du travail en France. Il est issu de Mihoubi (1994). Les données utilisées sont extraites de la base de données trimestrielle couvrant la période 1946-1989, construite par Laroque, Ralle, Salanié et Toujas-Bernate (1990) et portent sur l'ensemble de l'économie. Seules les séries de taux d'utilisation des capacités de production et de population active ne proviennent pas de cette base.

Le modèle comprend quatre équations.

- Une équation d'offre de travail :

$$\Delta \text{Ln}(L_t^s) = a_0 \Delta \text{Ln}(L_{t-1}^s) + a_1 \Delta \text{Ln}(W_{t-1} \cdot (1 - T_{S_{t-1}}) / P_{t-1}) + a_2 \Delta \text{Ln}(W_t^r / P_t) \\ + a_3 \Delta \text{Ln}(L_{t-1}^s / L_{t-1}) + a_4 \Delta \text{Ln}(L_{t-2}^s / POP_{t-2}) + a_5 \Delta \text{Ln}(W_{t-2} \cdot (1 - T_{S_{t-2}}) / P_{t-2}) \\ + a_6 \Delta \text{Ln}(W_{t-2}^r - P_{t-2}) + a_7 + \sigma_s u_{it},$$

- une équation de demande de travail :

$$\Delta \text{Ln}(L_t^d) = b_0 \Delta \text{Ln}(L_{t-1}^d) + b_1 \Delta \text{Ln}(K_t) + b_2 \Delta \text{Ln}(W_t \cdot (1 + T_{E_t}) / P_t) \\ + b_3 \Delta \text{Ln}(L_{t-1}^d / L_{t-1}) + b_4 \Delta \text{Ln}(L_{t-2}^d / K_{t-2}) + b_5 \Delta \text{Ln}(W_{t-2} \cdot (1 + T_{E_{t-2}}) / P_{t-2}) \\ + b_7 + \sigma_d v_{it},$$

¹²La maximisation de la pseudo-log-vraisemblance requiert la connaissance des dérivées des deux premiers moments par rapport aux paramètres. Les approximations de ces dérivées par la méthode de Monte-Carlo sont présentées dans l'article de Ben Jelili et Mihoubi (1992).

- une équation d'emploi effectif :

$$\ln L_t = \min(\ln(L_t^s), \ln(L_t^d));$$

- et une équation de salaire :

$$\begin{aligned} \Delta \ln W_t = & c_0 \Delta \ln W_{t-1} + c_1 \Delta \ln P_{c_t} + c_2 \Delta \ln P_{c_{t-1}} + c_3 \Delta \ln P_{c_{t-2}} + c_4 \Delta \ln (L_t^s / L_t^d) \\ & + c_5 \Delta \ln (L_{t-1}^s / L_{t-1}^d) + c_6 \ln (L_{t-2}^s / L_{t-2}^d) + c_7 + \sigma_w \omega_t \end{aligned}$$

avec

L : l'activité salariée;

T_s : le taux de cotisation salarié;

P : le prix du PIB;

P_c : le prix de consommation;

W^r : le salaire de réservation (celui-ci est approximé par le ratio des revenus non salariaux nets de l'épargne à la population active au sens démographique);

W : le salaire horaire brut;

POP : l'activité potentielle (produit de la population active et du nombre d'heures hebdomadaires travaillées);

K : le capital;

T_E : le taux de cotisation employeur;

u_t, v_t et ω_t suivent une loi N(0,1).

Pour obtenir des valeurs initiales raisonnables des différents paramètres, nous avons estimé indépendamment les équations d'offre et de demande, par les Moindres Carrés ordinaires (MCO), sur les sous-périodes où chacun des deux régimes semble *a priori* prédominer (de 1961 à 1973 pour l'équation d'offre, de 1975 à 1989 pour l'équation de demande).

Les résultats de l'estimation du modèle sont :

- pour l'équation de demande :

$$\begin{aligned} \Delta L_n (L_t^d) &= 0.23 \Delta L_n (L_{t-1}^d) + 1.04 \Delta L_n K_t - 0.27 \Delta L_n (W_t \cdot (1+T_{E_t}) / P_t) \\ &+ 0.67 \Delta L_n (L_{t-1}^d / L_{t-1}) - 0.05 \Delta L_n (L_{t-2}^d / K_{t-2}) - 0.06 \Delta L_n (W_{t-2} \cdot L_n(1+T_{E_{t-2}}) / P_{t-2}) \\ &- 0.57 + 0.0024 v_t, \end{aligned}$$

• pour l'équation d'offre :

$$\begin{aligned} \Delta L_n (L_t^s) &= 0.30 \Delta L_n (L_{t-1}^s) + 0.22 \Delta L_n (W_{t-1} \cdot (1-T_{S_t}) / P_{C_{t-1}}) - 0.10 \Delta L_n (W_t^r / P_{C_{t-1}}) \\ &+ 0.25 \Delta L_n (L_{t-1}^s / L_{t-1}) - 0.41 \Delta L_n (L_{t-2}^s / POP_{t-2}) + 0.094 \Delta L_n (W_{t-2} \cdot (1-T_{S_{t-2}}) / P_{C_{t-2}}) \\ &- 0.049 (W_{t-2}^r / P_{C_{t-2}}) + 0.53 + 0.002 u_t, \end{aligned}$$

• et pour l'équation de salaire :

$$\begin{aligned} \Delta L_n W_t &= 0.74 \Delta L_n (W_t / W_{t-1}) + 0.38 \Delta L_n P_{C_t} - 0.03 \Delta L_n P_{C_{t-1}} - 0.08 \Delta L_n P_{C_{t-2}} - 0.16 \Delta L_n (L_t^s / L_t^d) \\ &+ 0.14 \Delta L_n (L_{t-1}^s / L_{t-1}^d) - 0.008 \Delta L_n (L_{t-2}^s / L_{t-2}^d) + 0.002 + 0.0038 \omega_t \end{aligned}$$

Plusieurs enseignements originaux avec un modèle de déséquilibre peuvent être tirés .

D'une part, la demande de travail serait sensible aux variations des rationnements passés ($\Delta L_n (L_{t-1}^d / L_{t-1})$). L'effet de reprot dynamique, dont l'ampleur est voisine de 0.7, semble robuste aux spécifications du salaire de réservation. En revanche, l'effet de report dynamique est sensiblement moins robuste pour l'équation d'offre de travail (le signe et l'ampleur du coefficient lui étant associé varient suivant la définition du salaire de réservation).

D'autre part, les salaires seraient sensibles à la fois aux déséquilibres présents et passés. En outre, les variations et le niveau des déséquilibres peseraient sur la formation des salaires.

Enfin, la croissance des salaires est indexée unitairement sur la croissance du prix de consommation à très court terme (les termes d'inflation retardés sont en général non significatifs)

La spécification peut gagner en réalisme en relâchant l'hypothèse d'homogénéité du marché agrégé du travail, ce qui revient à prendre en compte l'agrégation de micro marchés en déséquilibre. On opte pour une modélisation de la dispersion des micro marchés répondant au soucis de simplicité et permettant de rendre compte d'une évolution tendancielle :

$$\tau = \tau_0 + \tau_1 t$$

Les résultats obtenus sont :

- pour l'équation de demande :

$$\begin{aligned} \Delta \ln (L_t^d) = & 0.46 \Delta \ln (L_{t-1}^d) + 0.88 \Delta \ln K_t - 0.15 \Delta \ln (W_t \cdot (1+T_{Et}) / P_t) \\ & + 0.61 \Delta \ln (L_{t-1}^d / L_{t-1}) - 0.04 \Delta \ln (L_{t-2}^d / K_{t-2}) - 0.04 \Delta \ln (W_{t-2} \cdot (1+T_{Et-2}) / P_{t-2}) \\ & - 0.37 + 0.0020 v_t, \end{aligned}$$

- pour l'équation d'offre :

$$\begin{aligned} \Delta \ln (L_t^s) = & 0.23 \Delta \ln (L_{t-1}^s) - 0.20 \Delta \ln (W_{t-1} \cdot (1-T_{St-1}) / P_{ct-1}) - 0.11 \Delta \ln (W_t^r / P_{ct-1}) \\ & + 0.28 \Delta \ln (L_{t-1}^s / L_{t-1}) - 0.24 \Delta \ln (L_{t-2}^s / POP_{t-2}) + 0.063 \Delta \ln (W_{t-2} \cdot (1-T_{St-2}) / P_{ct-2}) \\ & - 0.024 \Delta \ln (W_{t-2}^r / P_{ct-2}) + 0.39 + 0.0025 u_t, \end{aligned}$$

$$\tau = 0.027 t + 0.00$$

- et pour l'équation de salaire :

$$\begin{aligned} \Delta \ln W_t = & 0.82 \Delta \ln W_t + 0.42 \Delta \ln P_{ct} + 0.02 \Delta \ln P_{ct-1} - 0.22 \Delta \ln P_{ct-2} - 0.69 \Delta \ln (L_t^s / L_t^d) \\ & + 0.59 \Delta \ln (L_{t-1}^s / L_{t-1}^d) - 0.01 \Delta \ln (L_{t-2}^s / L_{t-2}^d) + 0.001 + 0.0027 \omega_t \end{aligned}$$

L'hypothèse de présence de micro marchés ne semble pas être rejetée, sans pour autant modifier fondamentalement les résultats obtenus précédemment.

Cet exemple illustre bien l'intérêt de la méthode du PMV pour estimer un modèle de déséquilibre aussi complexe. Celui-ci, riche d'enseignements économiques, combine à la fois la présence de micro marchés (un problème d'agrégation) et de variables endogènes (observées et latentes) retardées (analyse des comportements inter-temporels).

Annexe 1: Définitions et propriétés des estimateurs du PMV

A- Les hypothèses de White :

H1 : Les variables aléatoires $u_t, t = 1, \dots, T$, sont indépendamment distribuées et suivent une même loi f_0 ;

H2 : l'ensemble des paramètres Θ est compact ;

H3 : f est mesurable et continue quelque soit les valeurs de $\theta \in \Theta$. En plus f est intégrable par rapport à la vraie loi.

H4 : $E_0(f_0(u_t))$ existe et $|Ln(f(u_t, \theta))| \leq g(u_t), \forall \theta \in \Theta$, avec g une fonction intégrable par rapport à la vraie distribution de u_t ;

H5 : l'information de Kullback possède un minimum unique en $\theta = \theta^*$.

H6 : f est différentiable par rapport à $\theta \in \Theta$;

H7 : f est deux fois continuellement différentiable par rapport à $\theta \in \Theta$;

H8 : (i) θ^* est un point intérieur de Θ ;

(ii) la matrice $J(\theta) = E_x E_0 \left[\frac{\partial^2 Ln(f(u, \theta))}{\partial \theta \partial \theta} \right]$ existe et est inversible ;

(iii) la matrice $I(\theta) = E_x E_0 \left[\frac{\partial Ln(f(u, \theta))}{\partial \theta} \cdot \frac{\partial Ln(f(u, \theta))}{\partial \theta} \right]$ n'est pas singulière.

B- Les propriétés des familles des lois exponentielles linéaires et quadratiques

1- Famille de lois exponentielles linéaires

Gouriéroux, Monfort et Trognon (1984) détaillent quatre propriétés de ces familles :

Propriété 1 : $\frac{\partial A(m)}{\partial m} + \frac{\partial C(m)}{\partial m} m = 0$;

Propriété 2 : $\frac{\partial^2 A(m)}{\partial m \partial m} + \sum_{g=1}^G \frac{\partial^2 C_g(m)}{\partial m \partial m} m_g + \frac{\partial C(m)}{\partial m} = 0$ où C_g et m_g sont respectivement la gème composante de C et m ;

Propriété 3 : $\frac{\partial C(m)}{\partial m} = \Sigma^{-1}$ où Σ est la matrice de variance-covariance associée à $f(y,m)$. Cest une matrice définie positive de dimension G .

Propriété 4 : $A(m)+C(m)m_0 \leq A(m_0)+C(m_0)m_0, \forall m, m_0 \in M$.

2- Les familles exponentielles quadratiques

Une famille de mesure de probabilité sur \mathfrak{R}^G indicée par un paramètre $m \in M \subset \mathfrak{R}^G$ et $\Sigma(G \times G)$ (une matrice carrée définie positive) appartenant à un sous ensemble E des matrices définies positives est appelée famille exponentielle linéaire si :

- chaque élément de la famille a une fonction de densité par rapport à une mesure donnée et si cette fonction peut être écrite comme : $f(y,m,\Sigma) = \exp\{A(m,\Sigma) + B(y) + C(m,\Sigma)y + y'D(m,\Sigma)y\}$, $y \in \mathfrak{R}^G$ où $A(m,\Sigma)$ et $B(y)$ sont des scalaires, $C(m,\Sigma)$ est un vecteur ligne de dimension G et $D(m,\Sigma)$ représente une matrice carrée (G,G) ;

- m et Σ désignent respectivement la moyenne et la matrice de variance-covariance des aléas dont la distribution est $f(y,m,\Sigma)$.

Propriété 5 :

$A(m,\Sigma) + C(m,\Sigma)m_0 + \text{tr}D(m,\Sigma)(\Sigma_0 + m_0 m_0') \leq A(m_0,\Sigma_0) + C(m_0,\Sigma_0)m_0 + \text{tr}D(m_0,\Sigma_0)(\Sigma_0 + m_0 m_0')$,
 $\forall m, m_0 \in M$ et $\forall \Sigma, \Sigma_0 \in E$.

Propriété 6 : Pour $\alpha' = (m, \Sigma)$, nous vérifions :

- $E_0(y) = \int y f(y,m,\Sigma) v dy = m$;
- $E_0(y_2) = \int y_2^2 f(y,m,\Sigma) v dy = m_2 + \Sigma$;
- $\frac{\partial E_0(y)}{\partial \alpha} = \frac{\partial A}{\partial \alpha} m + \frac{\partial C}{\partial \alpha} (m_2 + \Sigma) + \frac{\partial D}{\partial \alpha} m_3$;
- $\frac{\partial E_0(y_2)}{\partial \alpha} = \frac{\partial A}{\partial \alpha} (m_2 + \Sigma) + \frac{\partial C}{\partial \alpha} m_3 + \frac{\partial D}{\partial \alpha} m_4$ où $m_j = E_0(y_j)$, $j=3,4$.

C- Les expressions des matrices I et J pour PMV2

$$I = E_x E_0 \left[\frac{\partial \text{Lnf}(y, m(x, \theta_0), g(x, \sigma_0))}{\partial p} \frac{\partial \text{Lnf}(y, m(x, \theta_0), g(x, \sigma_0))}{\partial p'} \right],$$

$$= E_x \left[\frac{\partial \alpha}{\partial p} \left(\frac{\partial C}{\partial \alpha} F + \frac{\partial D}{\partial \alpha} G \right) \frac{\partial \alpha'}{\partial p} \right]_{\theta=\theta_0, \sigma=\sigma_0},$$

avec :

$$F = \frac{\partial A}{\partial \alpha} m(x, \theta) + \frac{\partial C}{\partial \alpha} (m_2(x, \theta) + g(x, \sigma)) + \frac{\partial D}{\partial \alpha} m_3(x, \theta),$$

$$G = \frac{\partial A}{\partial \alpha} (m(x, \theta) + g(x, \sigma)) + \frac{\partial C}{\partial \alpha} m_3(x, \theta) + \frac{\partial D}{\partial \alpha} m_4(x, \theta).$$

$$J = -E_x E_0 \left[\frac{\partial^2 \text{Lnf}(y, m(x, \theta_0), g(x, \sigma_0))}{\partial p \partial p'} \right]$$

$$= -E_x \left[\frac{\partial \alpha}{\partial p} \left(\frac{\partial^2 A}{\partial \alpha \partial \alpha} + \frac{\partial^2 C}{\partial \alpha \partial \alpha} m(x, \theta) + \text{tr} \frac{\partial^2 D}{\partial \alpha \partial \alpha} [m_2(x, \theta) + g(x, \sigma)] \right) \frac{\partial \alpha'}{\partial p} \right]_{\theta=\theta_0, \sigma=\sigma_0},$$

où $m^3(x, \theta)$ et $m^4(x, \theta)$ représentent respectivement les moments d'ordre 3 et 4, $p'=(\theta, \sigma)$ et $\alpha'=(m(x, \theta), g(x, \sigma))$.

Annexe 2: Les méthodes traditionnelles d'estimation et l'agrégation de micro marchés en déséquilibre

A- Estimation par le Maximum de vraisemblance du modèle canonique de déséquilibre

Rappelons brièvement le principe du calcul de la vraisemblance dans le cadre du modèle canonique de déséquilibre. La densité de la transaction réalisée y_t à la date t est définie par :

$$h_t(y_t) = f_t(y_t / y_{dt}^* \geq y_{st}^*) \Pr(y_{dt}^* \geq y_{st}^*) + f_t(y_t / y_{dt}^* < y_{st}^*) \Pr(y_{dt}^* < y_{st}^*),$$

où f_t est la densité conditionnelle de la transaction réalisée sachant qu'on se situe dans un régime d'offre ($y_{dt}^* \geq y_{st}^*$) ou dans un régime de demande ($y_{dt}^* < y_{st}^*$). Cette densité est pondérée par la probabilité $\Pr(\cdot)$ d'être dans un des deux régimes.

Si $g_t(y_t, y_{st}^*)$ et $g_t(y_t, y_{st}^* / y_{dt}^* \geq y_{st}^*)$ désignent respectivement la densité jointe et la densité jointe conditionnelle de y_t , alors :

$$\begin{aligned} f_t(y_t / y_{dt}^* \geq y_{st}^*) &= \int_{y_t}^{+\infty} g_t(y_t, y_{dt}^* / y_{dt}^* \geq y_{st}^*) dy_{dt}^*, \\ &= \frac{\int_{y_t}^{+\infty} g_t(y_t, y_{dt}^*) dy_{dt}^*}{\Pr(y_{dt}^* \geq y_{st}^*)}. \end{aligned}$$

De la même façon, on définit :

$$\begin{aligned} f_t(y_t / y_{dt}^* < y_{st}^*) &= \int_{y_t}^{+\infty} g_t(y_t, y_{st}^* / y_{dt}^* < y_{st}^*) dy_{st}^*, \\ &= \frac{\int_{y_t}^{+\infty} g_t(y_t, y_{st}^*) dy_{st}^*}{\Pr(y_{dt}^* < y_{st}^*)}. \end{aligned}$$

Nous pouvons alors réécrire la densité de y_t :

$$h(y_t^r) = \int_{y_t^r}^{+\infty} g_t(y_{dt}^*, y_t^r) dy_{dt}^* + \int_{y_t^r}^{+\infty} g_t(y_{st}^*, y_t^r) dy_{st}^*,$$

où y_t^r désigne la réalisation de y_t .

Sous l'hypothèse d'indépendance des y_t , la vraisemblance (L) du modèle est égale au produit des vraisemblances de chaque période, d'où :

$$L = \prod_{t=1}^T h(y_t) \text{ avec } h_t \neq h_{t'} \text{ si } t \neq t'.$$

Lorsque les aléas sont normaux et indépendants, la densité de y_t s'écrit sous la forme :

$$h(y_t) = \frac{1}{\sigma_d} \varphi\left(\frac{y_t - x_{dt}\theta_d}{\sigma_d}\right) \Phi\left(\frac{x_{st}\theta_s - y_t}{\sigma_s}\right) + \frac{1}{\sigma_s} \varphi\left(\frac{y_t - x_{st}\theta_s}{\sigma_s}\right) \Phi\left(\frac{x_{dt}\theta_d - y_t}{\sigma_d}\right),$$

où Φ et φ sont respectivement la fonction de répartition et la densité de la loi normale centrée et réduite.

Les conditions nécessaires à la maximisation de la vraisemblance (ou de la log-vraisemblance) du modèle canonique conduisent à un système fortement non linéaire par rapport aux paramètres inconnus. Le recours à des méthodes itératives de résolution est donc nécessaire : à partir de valeurs initiales convenablement choisies pour les paramètres, l'algorithme d'optimisation évalue à chaque itération la fonction de vraisemblance. Les expressions analytiques des dérivées premières et secondes permettent de converger plus ou moins rapidement vers la solution qui maximise la fonction de vraisemblance.

L'estimateur du maximum de vraisemblance est convergent et asymptotiquement normal, mais malheureusement, ses propriétés numériques sont très mauvaises. En effet plusieurs types d'**extrema fallacieux** sont envisageables :

- Si l'un (ou les deux) des écarts-type (σ_i , $i = d, s$) tend vers zéro, la vraisemblance n'est plus bornée et tend alors vers l'infini. L'algorithme d'optimisation semble converger, mais vers une valeur nulle de l'écart-type. Hartley et Mallela (1977) montrent, cependant, que les estimateurs du maximum de vraisemblance sont convergents si le vecteur des paramètres vrais est choisi à l'intérieur d'un sous-espace compact de l'ensemble de tous les paramètres possibles, ce sous-espace ne comprenant pas les régions où $\sigma_s = 0$ et/ou $\sigma_d = 0$.

- L'un des deux régimes (excès d'offre, excès de demande) est affecté d'une probabilité d'apparition nulle sur l'ensemble de l'échantillon. Une telle situation provient, généralement, d'un nombre d'observations insuffisant pour observer l'ensemble des régimes ou d'une mauvaise spécification du régime non observé.

- Une difficulté supplémentaire peut surgir dans le cas où les aléas des équations d'offre et de demande sont corrélés. Il s'agit de la possibilité de convergence de l'algorithme d'optimisation en un point où le coefficient de

corrélation est de module unitaire, c'est-à-dire en un point où la fonction de vraisemblance n'est pas définie (Goldfeld et Quandt (1978)).

B- Déséquilibre et agrégation de micro marchés

Le modèle canonique de déséquilibre présente des limites liées à sa nature dichotomique. L'une des critiques majeures que l'on peut adresser à l'endroit de ce type de modèles au niveau agrégé concerne l'hypothèse trop forte du passage de l'économie dans son ensemble d'un régime (d'excès d'offre) à un autre (excès de demande). Cette hypothèse apparaît comme peu vraisemblable dès lors qu'on reconnaît l'existence d'hétérogénéité caractérisant l'économie.

L'analyse des déséquilibres en présence de micro marchés permet de relâcher cette hypothèse et de tenir compte de la simultanéité des demandes excédentaires et des offres excédentaires.

On considère alors un ensemble de micro marchés qui ne sont pas forcément affectés du même déséquilibre, le micro marché pouvant se définir comme l'entité élémentaire où l'ensemble des intervenants se rencontrent et où sont échangés des biens homogènes. On suppose que l'économie consiste en un nombre élevé N de marchés élémentaires (micro marchés), indicés par i ($i = 1, \dots, N$), pour lesquels les variables exogènes et les paramètres sont identiques, mais qui sont affectés, en plus des aléas macro-économiques (u_{st} et u_{dt}), par des aléas locaux ou aléas spatiaux (v_{sit} et v_{dit}) spécifiques à chacun des micro marchés. Les aléas locaux sont indépendants entre eux et indépendants des aléas macro-économiques. Sous ces conditions, le modèle canonique de déséquilibre décrivant la transaction réalisée sur le marché i à la période t s'écrit :

$$\begin{cases} y_{sit}^* = x_{st}\theta_s + \sigma_s u_{st} + \tau_s v_{sit}, \\ y_{dit}^* = x_{dt}\theta_d + \sigma_d u_{dt} + \tau_d v_{dit}, \\ y_{it} = \min(y_{sit}^*, y_{dit}^*), \end{cases} \text{ et } \begin{cases} i=1, \dots, N \\ t=1, \dots, T \end{cases} .$$

Le modèle agrégé est donné par l'espérance, par rapport aux aléas locaux, des quantités échangées sur les micro marchés :

$$y_t = E_v(y_{it} / x_{st}, x_{dt}, u_{st}, u_{dt}) .$$

Le passage de l'observation élémentaire à l'entité agrégée se traduit par le choix d'une procédure d'agrégation que doit respecter les distributions empiriques observées sur les unités élémentaires.

Les travaux fondateurs de Muellbauer (1978) sur la technique d'agrégation par intégration permettent de dériver les fondements communs à l'ensemble des modèles d'agrégation.

L'agrégation par intégration (par ailleurs développée par Kooiman et Kloeck (1979)) est fondée sur l'analyse de la situation des micro marchés à partir de celle observée sur les marchés agrégés. On suppose alors que le nombre de micro marchés est très grand et que la taille de chaque micro-marché est suffisamment faible. Ainsi l'absence d'information quant au nombre de marchés élémentaires est comblée par l'hypothèse d'un *continuum* de micro marchés. Grâce à cette hypothèse, la loi discrète suivie par les micro-offres et les micro-demandes peut être approchée par une distribution continue. le choix de la fonction de densité jointe s'avère crucial et c'est d'ailleurs en partie autour de ce choix que se distinguent les différentes méthodes d'agrégation. Ainsi les approches de Lambert et de Kooiman-Kloeck se distinguent essentiellement par les approximations conduisant à la transaction agrégée.

Lambert montre, qu'à la suite d'une série d'approximations¹³ et moyennant des hypothèses sur la distribution des aléas¹⁴, on aboutit à une forme CES de la transaction agrégée :

$$(5) \quad y = \left((x_{st}\theta_s)^{-\rho} + (x_{dt}\theta_d)^{-\rho} \right)^{-1/\rho}.$$

Le paramètre ρ est inversement proportionnel à l'indicateur de dispersion mesurant le déséquilibre entre l'offre et la demande sur les micro marchés (cet indicateur dépend des aléas locaux τ_s et τ_d). Cela distingue fondamentalement cette approche de celle de Ginsburgh, Tishler et Zang où l'approximation CES est totalement ad hoc.

L'approche de Lambert intègre l'existence du déséquilibre au niveau micro-économique. Elle est simple à appliquer¹⁵. Toutefois, elle repose sur des hypothèses assez fortes qui limitent sa portée. En outre, l'utilisation de données auxiliaires (données d'enquêtes) pour calculer le paramètre ρ (ce qui rend la procédure

¹³Les diverses approximations et hypothèses sont explicitées en annexe de la thèse de Lambert (1984).

¹⁴L'auteur retient l'hypothèse d'une distribution log-normale jointe de la demande et de l'offre sur les micro-marchés.

¹⁵Comme le paramètre ρ est directement relié à la proportion de micro-marchés en régime d'offre ou en régime de demande, on peut le calculer dès lors qu'un indicateur de cette proportion (on pense plus particulièrement aux données d'enquêtes) est disponible.

d'estimation encore plus simple) peuvent introduire des biais dans les estimations des paramètres¹⁶.

Dans le cas de l'approche développée par Kooiman-Kloek, la dérivation entraîne l'écriture de la transaction agrégée sous la forme d'une moyenne pondérée de l'offre et de la demande augmentée d'un terme de friction.

¹⁶Voir à ce sujet Kooiman (1984,1986).

Références bibliographiques

Ben Jelili, R., Mihoubi F. [1993] : "Estimation des modèles de déséquilibre dynamiques : une application de la méthode du pseudo-maximum de vraisemblance." *Document de travail* 9304 du LATEC, Dijon.

Burguete, J., Gallant et Souza, G. [1982] : "On unification of the asymptotic theory of nonlinear econometric models." *Econometric Review*, 1, pp.151-190.

Danielson, J. et Richard, J.F. [1992] : "Likelihood evaluation for dynamic latent variables models : control variates." *Document ronéotypé*.

Fair, R.C et Jaffee, D.M. [1972] : "Methods of estimation for markets in disequilibrium." *Econometrica*, 40, pp.497-514.

Goldfeld, S.M., Quandt, R.E. [1978] : "Some properties of the simple disequilibrium model with covariance." *Economics Letters*, 1, pp.343-346.

Gouriéroux, C. [1989] : *Econométrie des variables qualitatives*. *Economica*, Paris.

Gouriéroux, C. et Laroque ,G. [1985] : "The aggregation of commodities in quantity rationing models." *International Economic Review*, 26, pp.681-699.

Gouriéroux, C. et Monfort, A. et Renault, E. [1993] : "Indirect Inference.", *Journal of Applied Econometrics*, Special Issue, vol 8, December, pp.85-118.

Gouriéroux ,C., Monfort, A. et Trognon, A. [1984] : "Pseudo-maximum likelihood methods : theory." *Econometrica*, 52, p.681-700.

Hajivasiliou, V.A. et McFadden, D. [1989] : "Country heterogeneity and external debt crises : estimation by the method of simulated scores." *Document ronéotypé*.

Kooiman, P. et Kloek, T. [1979] : "aggregate of micro-markets in disequilibrium : theory and application to the Dutch labor market 1948-1975." *Document de travail*, Erasmus University.

Laffargue, J.P. et Mihoubi, F. [1989] : "Estimation en déséquilibre avec et sans micro marchés des équations du commerce extérieur de l'industrie par le pseudo-

maximum de vraisemblance." *Rapport au commissariat général du plan*, document ronéotypé.

Lambert, J.P. [1984] : "Disequilibrium macro models based on business survey data : theory and estimation for Belgian manufacturing sector." *PhD thesis*, Université Catholique de Louvain.

Laroque, G. et Salanié, B. [1989] : "Estimation of multi-market fix-price models : an application of pseudo-maximum likelihood methods." *Econometrica*, 57, pp.831-860.

Laroque, G. et Salanié, B. [1990] : "Estimating the canonical Disequilibrium model : asymptotic theory and finite sample properties." *Document de travail* 9005, INSEE.

Laroque, G. et Salanié, B. [1993] : "Simulated-based estimation of models with lagged latent variables." *Journal of Applied Econometrics*, Special Issue, vol 8, December, pp.119-133.

Laroque, G., Ralle, P., Salanié, B. et Toujas-Bernate, J. (1990) : "Description d'une base de données trimestrielles longues (1946:1 à 1989:4)." *Document de travail*, INSEE.

Lerman, S. et Manski, C. [1981] : "On the use of simulated frequencies to approximate choice probabilities." in *Manski, C. et McFadden, D. éditeurs, Structural analysis of discrete data with econometric applications*, Cambridge, MIT Press.

Maddala, G.S. et Nelson, F.D. [1974] : "Maximum likelihood methods for models of markets in disequilibrium." *Econometrica*, 42, pp.1013-1030.

McFadden, D. [1989] : "A method of simulated moment for estimation of discrete response models without numerical integration." *Econometrica*, 57, pp.995-1026.

Mihoubi, F. [1994] : "Estimation, tests et évaluation du modèle de déséquilibre avec micro marchés par le pseudo-maximum de vraisemblance : application à l'emploi en France." *Thèse de Doctorat* de l'Université de Paris I.

Muellbauer, J. [1978] : "Macrotheory vs. Macroeconometrics : the treatment of disequilibrium in macromodels." *Document de travail* 59, Department of economics, Birbeck College, University of London.

Pakes, A. et Pollard, D. [1989] : "Simulation and the asymptotics of optimisation estimators." *Econometrica*, 57, pp.1027-1057.

Quandt, R.E. [1982] : "Econometric disequilibrium models." *Document de travail* 491, CORE, Louvain-la-Neuve.

Quandt, R.E. [1988] : "The econometrics of disequilibrium." *Basil Blackwell Ed.*

White, H. [1982] : "Maximum likelihood estimation of misspecified models." *Econometrica*, 50, pp.1-25.