ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 62 – 2001

Endogénéité d'une variable explicative dichotomique dans le cadre d'un modèle probit bivarié

Une application au lien entre fécondité et activité féminine

Stéfan LOLLIVIER *

RÉSUMÉ. – L'endogénéité des variables explicatives, due à une corrélation entre celles-ci et le terme d'erreur d'une régression, peut entraîner des biais importants dans l'estimation des comportements. Lorsque la variable expliquée et les variables explicatives sont continues, on dispose de plusieurs méthodes pour détecter l'endogénéité et corriger les biais. En revanche, lorsque les deux types de variables sont dichotomiques, la résolution des problèmes liés à l'endogénéité est plus délicate et rarement effectuée.

On présente ici un test d'endogénéité pouvant être mis en œuvre très simplement à partir des logiciels usuels ainsi qu'une méthode d'estimation ayant recours à des techniques d'intégration numériques classiques. Les deux procédures sont ensuite appliquées à la détection de l'endogénéité d'une variable de fécondité dans un modèle de choix d'activité chez les femmes dont le conjoint est salarié. Les deux techniques concluent de façon très similaire à une absence d'endogénéité.

Endogeneity of a Dummy Regressor in a Bivariate Probit Model

An application to the Link between Fertility and Female Activity

ABSTRACT. – The endogeneity of regressors, due to a correlation with the error term, can involve significant biases in the estimation of behaviors. When both the explained variable and the regressors are continuous, several methods to detect the endogeneity and correct its effects are available. However, when the two variables are dichotomic, the resolution of the problems due to the endogeneity is more difficult and seldom carried out.

A test of endogeneity, which can be computed very simply with the usual software, is presented in the paper. Moreover, a method of estimation based on numerical techniques of integration is developed. The two procedures are then applied to detect the endogeneity of fertility in a model of female activity when the husband is not self-employed. The two techniques conclude in a very similar way to an absence of endogeneity.

---

* S. LOLLIVIER : ENSAE.

L'auteur remercie deux rapporteurs anonymes pour leurs remarques et propositions. Les erreurs ou omissions qui pourraient subsister sont de la seule responsabilité de l'auteur.

1 Introduction

L'endogénéité des variables explicatives pose souvent problème lorsque l'on s'intéresse à l'économétrie des comportements. En présence d'endogénéité, l'espérance du terme d'erreur conditionnelle à la variable explicative n'est plus nulle et les estimateurs habituels présentent des biais. Cependant, le traitement de l'endogénéité des variables explicatives est maintenant bien établi lorsque les variables expliquée et explicative sont continues. La méthode des régressions augmentées permet de réaliser à moindres frais des tests de spécification mais aussi d'effectuer des estimations permettant de corriger de tels biais (voir, par exemple, ROBIN [2000]). On dispose ainsi de procédures efficaces pour traiter ces biais, la difficulté dans la pratique consistant seulement à trouver des instruments adéquats.

Cependant, de plus en plus de travaux appliqués ont recours à des estimations économétriques faisant appel à des variables qualitatives. Celles-ci peuvent apparaître en tant que variables explicatives ou bien expliquées. La résolution du modèle simultané est alors beaucoup plus délicate. HECKMAN [1978] a présenté une formalisation générale permettant de traiter le problème lorsque la variable expliquée est continue et la variable explicative dichotomique. Comme pour la correction du biais de sélectivité dans le modèle Tobit généralisé, il montre qu'il n'est pas absolument nécessaire d'avoir recours à la méthode du maximum de vraisemblance sur le système bivarié, mais que des procédures en deux étapes sont envisageables. Il est, en particulier, possible d'avoir recours à une régression linéaire augmentée d'un terme issu d'une première étape d'estimation du modèle probit conditionnant la variable qualitative. L'estimateur des moindres carrés ordinaires portant sur cette régression est alors convergent (HECKMAN [1974]). La difficulté consiste à réaliser le test permettant de conclure ou non à l'endogénéité, la matrice de variance-covariance des estimateurs n'étant pas celle qui ressort de l'estimation des moindres carrés ordinaires. LEE, MADDALA et TROST [1980] ont en effet montré que la matrice de variance-covariance correcte présentait une forme complexe, faisant intervenir non seulement la correction d'hétéroscédasticité, mais aussi un terme provenant du fait que le régresseur complémentaire était issu d'une estimation et non d'une observation. Dans la pratique, le calcul est assez lourd, et il est plus commode de mettre en œuvre la méthode préconisée par WHITE [1980], qui permet de calculer de façon convergente la matrice de variance-covariance à partir des résidus observés.

La complexité s'accroît encore d'un degré lorsque la variable expliquée est elle aussi qualitative. Même si les développements classiques restent vrais sur les variables latentes sous-jacentes aux variables dichotomiques, les problèmes liés à l'estimation se posent dès lors que l'on n'a plus affaire à des densités mais à des probabilités, ce qui entraîne de nombreuses complexités numériques. De fait, le problème a été relativement peu exploré dans la littérature. Dans ce papier, on se propose précisément :

– de développer un test simple d'endogénéité utilisant la méthode des tests du score sur régressions auxiliaires. Le recours à de telles méthodes, exposées dans DAVIDSON et MAC KINNON [1990] et MAC KINNON [1992], permet de réaliser un test portant sur une pseudo-régression dont les termes sont calculés à partir des gradients de la log-vraisemblance évalués sous l'hypothèse nulle. Relativement facile à mettre en œuvre, une telle technique ne nécessite que le recours à l'estimation du modèle sous l'hypothèse nulle et évite l'estimation sous l'hypothèse alternative. Des tests de cette nature sont fréquemment utilisés comme tests de spécification dans les modèles logit ou probit (BALTAGI [1999]).

– de réaliser l'estimation proprement dite du maximum de vraisemblance sur le modèle normal bivarié. La difficulté est alors d'ordre purement numérique, puisqu'il s'agit de calculer une intégrale double. MROZ [1999], dans un contexte un peu différent, suggère un conditionnement permettant de se ramener à une intégration simple avec une densité normale. Des méthodes de quadrature numérique, analogues à celles développées par BUTLER et MOFFIT [1982], permettent, avec le recours aux formules d'intégration d'*Hermite*, d'obtenir une convergence avec des temps de calcul particulièrement brefs.

Ces deux techniques sont ensuite appliquées pour étudier le caractère endogène d'une variable de fécondité lorsque l'on cherche à évaluer son influence sur le comportement d'activité féminine. On reproche fréquemment aux travaux empiriques de négliger l'endogénéité de la fécondité dans les modèles de participation, ce qui est de nature à entraîner des biais dans l'estimation de l'impact de la fécondité sur l'activité. Dans l'étude, on utilise les données de la première vague du panel européen, en se restreignant classiquement aux femmes vivant en couple dont le conjoint est salarié. On se limite également aux femmes âgées de 35 ans et plus, la plupart ayant eu tous les enfants qu'elles désiraient. On évite ainsi de mélanger des populations de mères dont le nombre d'enfants est le même au moment de l'enquête, mais dont la descendance ultérieure pourra différer. Or, la descendance totale prévue joue au moins autant sur le comportement d'activité que la descendance courante, ce qui pose des problèmes d'interprétation (LOLLIVIER [2000]). La variable retraçant le comportement de fécondité est une variable indicatrice du fait d'avoir au moins trois enfants. On verra que dans l'échantillon retenu, ce seuil est le plus pertinent pour discriminer le comportement d'activité : les femmes sans enfant et les mères de moins de trois enfants ont quasiment le même comportement ; ce n'est qu'à partir du troisième que l'activité est moins intense.

Les deux méthodes s'accordent à rejeter l'hypothèse alternative d'endogénéité, si l'on spécifie correctement le modèle. Cette concordance est d'autant plus remarquable que les tests fondés sur les régressions auxiliaires sont fréquemment critiqués pour leurs mauvaises propriétés sur petit échantillon (DAVIDSON, MAC KINNON [1990]). Ici, l'échantillon dépasse les 1 850 femmes et on est sans doute plus proche des conditions asymptotiques que dans les simulations destinées à évaluer la performance de ces tests. Une évaluation complémentaire visant à mettre en œuvre les deux procédures sur un modèle mal spécifié conclue, cette fois, à l'endogénéité, de façon analogue pour les deux méthodes.

2 Le modèle probit bivarié

On considère le modèle probit bivarié :

$$\begin{cases} Y_1^* = X_1\beta_1 + \varepsilon_1 \\ Y_2^* = X_2\beta_2 + \alpha Y_1 + \varepsilon_2 \end{cases},$$

avec

$$\begin{pmatrix}\varepsilon_1 \\ \varepsilon_2\end{pmatrix} \to N\left[\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right].$$

Y_1^* et Y_2^* sont deux variables latentes, pour lesquelles on observe $Y_1 = I(Y_1^* > 0)$ et $Y_2 = I(Y_2^* > 0)$.

La loi conditionnelle de Y_2^* sachant Y_1^* peut s'écrire :

$$Y_2^* = X_2\beta_2 + \alpha Y_1 + \rho\frac{\sigma_2}{\sigma_1}(Y_1^* - X_1\beta_1) + u,$$

le terme d'erreur u suivant une loi normale d'espérance nulle et de variance $\sigma_2^2(1-\rho^2)$.

Dès lors que $\rho \neq 0$, on voit que $E(\varepsilon_2/X_2,Y_1) \neq 0$. L'estimation autonome de la deuxième équation du système initial peut comporter un biais d'endogénéité. Dans ce type de problème, on sait en outre que les paramètres ne peuvent être déterminés qu'à une constante multiplicative près. Afin de restaurer l'identifiabilité, on contraint fréquemment les termes d'erreur à avoir une variance unitaire (σ_1 et $\sigma_2 = 1$).

2.1 Construction d'un test simple d'endogénéité

Comme on le verra par la suite, l'estimation du modèle probit bivarié par la méthode du maximum de vraisemblance est relativement complexe, puisqu'elle nécessite le recours à une intégration numérique due au fait que l'aléa est bivarié. On peut cependant proposer un test d'endogénéité beaucoup plus simple à mettre en œuvre que les tests habituels de restriction (rapport de vraisemblance, Wald,…). Ces derniers nécessitent, en règle générale, l'estimation du modèle à la fois sous l'hypothèse nulle et sous l'hypothèse alternative. Le test du score proposé ici ne va requérir que l'estimation du modèle sous l'hypothèse nulle, c'est-à-dire en l'absence de corrélation entre les termes d'erreurs. Très concrètement, sous l'hypothèse nulle, le modèle se ramène à deux modèles probit indépendants. Leur estimation est intégrée dans tous les logiciels courants.

La méthode consiste à appliquer des tests du score en réalisant des ajustements sur des régressions auxiliaires (Davidson et Mac Kinnon [1990]). Elle est utilisée pour réaliser différents tests de spécification (Baltagi [1999]).

Soit un échantillon de taille $N (i = 1,\ldots,N)$ sur lequel on cherche à estimer un modèle conduisant à la log-vraisemblance :

$$L(\theta) = L(\theta_1,\theta_2) = \sum_{i=1}^{N} l_i(\theta_1,\theta_2).$$

θ_1 est un vecteur de paramètres de taille k_1 et θ_2 un vecteur de paramètres de taille k_2, avec $k = k_1 + k_2$.

On note $G(\theta)$ la matrice de taille (N,k) dont l'élément générique $G_{i,j}(\theta) = \dfrac{\partial l_i(\theta)}{\partial \theta_j}$ est égal au gradient de la contribution de l'individu i à la log-vraisemblance. Le vecteur $g(\theta)$, de taille $(k,1)$, correspondant au gradient de la log-vraisemblance calculée sur tout l'échantillon s'écrit alors :

$$g(\theta) = G(\theta)'e \ \text{ soit } \ g_j(\theta) = \sum_{i=1}^{N} G_{i,j}(\theta)$$

où e est un vecteur de taille N composé de 1.

Soit maintenant $\hat{\theta}$ un estimateur convergent de θ. La matrice d'information de *Fisher* $I(\theta)$ peut être estimée de façon convergente à partir du *Hessien* calculé en θ. Mais on sait que l'on dispose d'un autre estimateur convergent $\hat{I}$, que l'on peut construire plus aisément à partir des seules dérivées premières $G(\theta)$:

$$\hat{I} = G(\hat{\theta})'G(\hat{\theta})$$

Cet estimateur de la matrice d'information de *Fisher* est très fréquemment utilisé dans les algorithmes d'estimation du maximum de vraisemblance, notamment ceux de *Berndt*, *Hall*, *Hall* et *Hausmann*.

Dans le modèle probit bivarié, le test d'endogénéité correspond à un simple test de spécification correspondant à la restriction $\theta_2 = 0$ (ici $\rho = 0$). Soit $\tilde{\theta}_1$ l'estimateur du maximum de vraisemblance sous l'hypothèse nulle $\theta_2 = 0$, et notons $\tilde{\theta} = (\tilde{\theta}_1,0)$. La restriction peut être testée en utilisant la statistique du score sous l'hypothèse nulle :

$$LM = g(\tilde{\theta})'\tilde{I}^{-1}g(\tilde{\theta}),$$

formule dans laquelle $\tilde{I} = G(\tilde{\theta})'G(\tilde{\theta})$ est un estimateur convergent de la matrice d'information de *Fisher*. Cette statistique du score suit asymptotiquement un $\chi^2(k_2)$. Elle peut être réécrite :

$$LM = \left[e'G(\tilde{\theta})\right]\left[G(\tilde{\theta})'G(\tilde{\theta})\right]^{-1}\left[G(\tilde{\theta})'e\right]$$

Formulée ainsi, la statistique du score correspond exactement à la somme des carrés de la variable expliquée dans la régression auxiliaire :

$$e = G(\tilde{\theta})b + \text{erreur}$$

Cette régression consiste simplement à expliquer par les moindres carrés ordinaires un vecteur composé de 1 par les variables correspondant aux colonnes de la matrice G, calculée sous **l'hypothèse nulle**.

Dans le cas du modèle probit bivarié, avec $\sigma_1 = 1$ et $\sigma_2 = 1$, la contribution d'un individu à la log-vraisemblance s'écrit, si l'on omet l'indice de l'individu :

$$(1) \quad l = \text{Log}\left[\int_{D(Y_1)} Z_2\left(\frac{X_2\beta_2 + \alpha Y_1 + \rho(Y_1^* - X_1\beta_1)}{\sqrt{1-\rho^2}}\right)\varphi(Y_1^* - X_1\beta_1)dY_1^*\right]$$

où :

– $D(Y_1)$vaut $]-\infty,0]$ si $Y_1 = 0$ ou $]0,+\infty[$ si $Y_1 = 1$,
– φ est la densité de la loi normale unidimensionnelle,
– $Z_l(x) = \Phi(x)$ si $Y_l = 1$ et $Z_l(x) = 1 - \Phi(x)$, si $Y_l = 0$ $(l = 1,2)$,

Φ étant la fonction de répartition de la loi normale unidimensionnelle.

On en déduit la valeur du gradient sous l'hypothèse nulle $\rho = 0$ (voir démonstration en annexe) :

$$\begin{cases} \dfrac{\partial l}{\partial \beta_2} = \dfrac{\varphi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)(Y_2 - \Phi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1))}{\Phi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)(1 - \Phi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1))}X_2' \\ \dfrac{\partial l}{\partial \alpha} = \dfrac{\varphi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)(Y_2 - \Phi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1))}{\Phi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)(1 - \Phi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1))}Y_1 \\ \dfrac{\partial l}{\partial \beta_1} = \dfrac{\varphi(X_1\tilde{\beta}_1)(Y_1 - \Phi(X_1\tilde{\beta}_1))}{\Phi(X_1\tilde{\beta}_1)(1 - \Phi(X_1\tilde{\beta}_1))}X_1' \\ \dfrac{\partial l}{\partial \rho} = \dfrac{\varphi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)(Y_2 - \Phi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1))\varphi(X_1\tilde{\beta}_1)(Y_1 - \Phi(X_1\tilde{\beta}_1))}{\Phi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)(1 - \Phi(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1))\Phi(X_1\tilde{\beta}_1)(1 - \Phi(X_1\tilde{\beta}_1))} \end{cases}$$

où $(\tilde{\beta}_2,\tilde{\alpha},\tilde{\beta}_1,0)$ correspond à l'estimateur de $(\beta_2,\alpha,\beta_1,\rho)$ sous l'hypothèse nulle.

Il suffit ensuite d'ajuster par les moindres carrés ordinaires un vecteur $(N,1)$ composé de 1 sur l'ensemble de ces régresseurs, soit une matrice de taille $(N,\dim(\beta_2) + \dim(\beta_1) + 2)$, pour obtenir la statistique de test. Les logiciels usuels permettent d'estimer des modèles probit, et fournissent les valeurs des paramètres sous l'hypothèse nulle. Connaissant la valeur de $(\tilde{\beta}_2,\tilde{\alpha},\tilde{\beta}_1)$, les calculs des régresseurs puis de la statistique de test ne nécessitent pas le recours à une programmation spécifique.

On reproche fréquemment à ces tests leurs mauvaises propriétés à distance finie, surtout lorsque les hypothèses sur la loi du terme d'erreur ne sont pas vérifiées (GODFREY, MCALEER, MC KENZIE [1988]). En effet, sur petit échantillon, l'approximation de la matrice d'information de *Fisher* par $G(\theta)'G(\theta)$ est de mauvaise qualité, ce qui fait que la distribution du test correspond assez mal à la distribution asymptotique. En pratique, des simulations montrent que ces tests ont parfois tendance à rejeter l'hypothèse nulle de façon incorrecte. Cependant, l'échantillon de données individuelles auquel on va s'intéresser ici est sans doute de taille suffisante pour que les conditions soient proches des conditions asymptotiques. C'est du moins ce que va suggérer la comparaison des résultats obtenus dans cette partie avec ceux calculés avec des tests usuels.

2.2 **Estimation du modèle probit bivarié avec corrélation des perturbations**

Pour estimer le modèle probit bivarié, on reprend une approche introduite notamment par MROZ [1999], consistant à redéfinir les termes d'erreur dans le modèle initial :

$$\varepsilon_1 = \rho_1 \nu + u_1$$
$$\varepsilon_2 = \rho_1 \nu + u_2$$

où u_1, u_2 et ν sont supposés de moyenne nulle, indépendants entre eux et indépendants des variables explicatives du modèle. On peut interpréter cette formulation en considérant ν comme une variable aléatoire inobservée qui influence linéairement les deux perturbations. Dans l'étude, on supposera que les trois variables sont distribuées selon une loi normale et que ν a une variance unitaire. Cette hypothèse peut cependant être relâchée, et on peut notamment chercher à approximer la fonction de répartition de ν par une fonction en escalier (MROZ [1999]). L'estimation est alors beaucoup plus complexe et peut être effectuée en s'inspirant de la démarche itérative suggérée par HECKMAN et SINGER [1984].

Conditionnellement à la valeur prise par ν, la distribution jointe de ε_1 et ε_2 est donnée par la formule :

$$f(\varepsilon_1, \varepsilon_2/\nu) = \frac{1}{s_1}\varphi\left(\frac{\varepsilon_1 - \rho_1 \nu}{s_1}\right)\frac{1}{s_2}\varphi\left(\frac{\varepsilon_2 - \rho_2 \nu}{s_2}\right)$$

où s_1 et s_2 représentent les écart-types de u_1 et u_2 et φ la densité de la loi normale. Si on note Φ la fonction de répartition de la loi normale, la distribution jointe et non conditionnelle de ε_1 et ε_2 peut s'écrire :

$$f(\varepsilon_1, \varepsilon_2) = \int f(\varepsilon_1, \varepsilon_2/\nu) d\Phi(\nu)$$

Elle correspond à la distribution d'une loi normale bidimensionnelle, de moyenne nulle, de variances $(s_1^2 + \rho_1^2)$ et $(s_2^2 + \rho_2^2)$ et de covariance $\rho_1\rho_2$.

Le modèle ainsi redéfini n'est pas identifiable, mais il se présente sous la forme standard d'une intégration gaussienne. On sait qu'alors l'intégrale peut être approximée avec une très grande précision au moyen d'une somme en ayant recours à la formule d'intégration d'*Hermite* (BUTLER et MOFFIT [1982]) :

$$\int_{+\infty}^{+\infty} g(Z)\exp(-\nu^2)dZ = \sum_{g=1}^{G} \omega_g g(\nu_g)$$

où G est le nombre de points ν_g où la fonction est évaluée, et ω_g les poids des différents points. Les valeurs de ces paramètres sont disponibles dans différents ouvrages contenant des tables mathématiques (ABRAMOWITZ, STEGUN [1964]). Dans l'application, on a choisi $G = 9$, ce qui fournit une excellente approximation, au vu des nombreuses simulations réalisées sur le sujet.

Si l'on omet comme précédemment l'indice de l'individu, sa contribution à la vraisemblance s'écrit :

$$l = \mathrm{Log}\left[\sum_{g=1}^{G} \sqrt{2}\omega_g Z_2\left(\frac{X_2\beta_2 + \alpha Y_1 + \rho_2\sqrt{2}\nu_g}{s_2}\right) Z_1\left(\frac{X_1\beta_1 + \rho_1\sqrt{2}\nu_g}{s_1}\right)\right],$$

le terme $\sqrt{2}$ apparaissant du fait des différences entre la forme de la densité et la formule d'intégration précédente. Les fonctions Z_l ont la même signification que dans la formule (1). Pour assurer l'identification, on a en outre introduit les contraintes :

$$s_1 = \rho_1 = \frac{1}{\sqrt{2}}$$

$$s_2 = 1$$

La première restriction permet d'assurer $\sigma_1 = 1$ comme dans le modèle initial. Pour des raisons de simplification des calculs, on n'a pas respecté la même analogie pour σ_2. Sous cette forme, le calcul de la log-vraisemblance et de ses dérivées est relativement aisé, le recours au conditionnement ayant permis d'éviter le calcul d'une intégrale double.

3 Application numérique

3.1 L'échantillon

L'échantillon dont on dispose est constitué par la première vague du panel européen des ménages. Il s'agit d'une opération communautaire, coordonnée par EUROSTAT, dont l'objectif principal est de suivre les dynamiques d'emploi et de revenu. La première vague a eu lieu en 1994 dans les douze états formant l'Union européenne de l'époque. L'échantillon total comprenait environ 60 500 ménages sélectionnés de façon aléatoire avec un taux de sondage uniforme. Cette enquête permet notamment de fournir des statistiques entre pays afin d'éclairer les aspects sociaux de la politique communautaire. D'autres vagues ultérieures à 1994 ont été collectées, mais ne sont pas utilisées ici.

En France, l'INSEE assure la gestion et la collecte de l'information. La partie du questionnaire utilisée dans l'étude est celle relative aux individus qui vivent dans les ménages interrogés. Le questionnaire décrit la situation des individus face au marché du travail, mais aussi l'environnement familial dans lequel vivent les personnes. En outre, la première vague de 1994 recense tous les enfants issus des individus, notamment ceux qui vivent hors du ménage, et fournit des renseignements sur la jeunesse de l'individu, utiles pour les besoins de l'étude.

Parmi les femmes vivant en couple, on va se restreindre aux femmes âgées de 35 à 56 ans en 1994, et dont les conjoints sont salariés ou anciens salariés. La borne supérieure sur l'âge permet de se limiter aux périodes de vie active, sans que les problèmes de pré-retraite ne viennent trop perturber la notion d'activité. La borne inférieure permet de se restreindre aux femmes ayant eu la quasi-totalité de leurs enfants. L'autre restriction, concernant le statut du conjoint, est classique dans ce genre de travail. Elle permet d'exclure les femmes d'indépendants dont le statut éventuel d'aide familiale est de nature différente de la notion habituelle d'activité chez les salariés. En outre, le revenu des indépendants n'est pas toujours comparable à celui des salariés, du fait des règles fiscales et comptables en vigueur en France.

3.2 Les caractéristiques des femmes conjointes de salarié

Au total, l'échantillon se compose en 1994 de 1 850 femmes (tableau 1). Le taux d'activité moyen est de 71 % et diminue avec l'âge. L'activité est d'autant plus fréquente que la femme dispose d'un diplôme de niveau élevé. Ce résultat attendu s'explique dans le cadre du modèle habituel d'arbitrage entre travail et loisir (voir, par exemple, BOURGUIGNON [1986], DAGSVIK, LAISNEY, STROM, OSTERVOLD [1988] pour une description de modèles tenant compte de la fiscalité). Selon celui-ci, l'activité est d'autant plus fréquente que les revenus qu'elle procure sont élevés et que les coûts liés à l'activité sont faibles. On vérifie donc que le diplôme, qui permet d'accéder à des emplois mieux rémunérés, joue positivement sur l'arbitrage. En revanche, l'influence du revenu du conjoint est conforme à celle qui ressort de travaux analogues, mais à première vue plus surprenante. On s'attendrait à ce qu'un revenu du conjoint plus élevé décourage l'activité du fait de la progressivité du système fiscal : les prélèvements sur un même salaire féminin sont d'autant plus élevés que le salaire du mari est conséquent, puisque la tranche marginale d'imposition est plus grande, ce qui n'est pas le cas sur le tableau 1. En fait, le profil est perturbé par la forte homogamie sociale qui conduit les femmes diplômées à vivre avec des conjoints eux aussi diplômés. À diplôme équivalent, l'estimation de modèles probit simples montre un effet négatif du revenu du conjoint en haut de la distribution (tableau 3).

TABLEAU 1
Caractéristiques de l'échantillon

	Effectif	Taux d'activité	% de 3 enfants ou plus
Âge			
De 35 à 39 ans	480	75,6	34,6
De 40 à 44 ans	475	74,5	37,3
De 45 à 49 ans	477	71,5	33,1
De 50 à 54 ans	300	63,0	42,0
55 ans et plus	118	51,7	44,9
Diplôme			
Pas de diplôme	351	52,4	56,7
CEP	373	67,3	42,4
CAP-BEP	311	76,5	29,9
BEPC	301	73,4	29,9
Bac technique	75	86,7	20,0
Bac général	142	68,3	30,3
Supérieur au Bac	297	84,8	27,6
Salaire mensuel du conjoint (en F97)			
Conjoint chômeur	107	72,9	46,7
Moins de 5 800 F	95	61,1	50,5
De 5 800 à 7 100 F	170	63,5	44,1
De 7 100 à 9 500 F	422	72,0	39,1
De 9 500 à 13 300 F	426	75,1	32,2
De 13 300 à 19 800 F	296	75,0	28,0
Plus de 19 800 F	223	65,9	31,8
Non déclaré	111	64,0	45,9
Nombre d'enfants			
0	105	79,0	0
1	310	83,2	0
2	755	80,3	0
3 et plus	680	53,1	100
Née en France	1 584	72,1	34,1
Née hors de France	266	62,4	52,6
Mère inactive pendant l'enfance	1 052	65,8	39,8
Mère active pendant l'enfance	798	77,2	32,7
Profession du père pendant l'enfance			
Agriculteur	253	70,0	44,3
Indépendant	178	75,8	25,8
Cadre	140	73,6	36,4
Profession Intermédiaire	193	74,6	32,6
Employé	210	71,4	36,2
Ouvrier	711	67,4	38,1
Autre	165	72,7	37
Ensemble	1 850	70,7	36,8

Champ : Femmes conjoint de salariés âgées de 35 à 56 ans.
Source : Panel européen 1994, Insee.

Une des particularités de l'échantillon est de pouvoir disposer du nombre total d'enfants de la personne et non des seuls enfants présents au domicile du foyer parental. Le nombre d'enfants s'accroît en début de cycle de vie mais se stabilise à partir de 35 ans, ce qui justifie la restriction de champ effectuée dans l'étude (tableau 2). On vérifie enfin le fait qu'un nombre d'enfants élevé va de pair avec une activité moindre, ce qui est qualitativement conforme avec l'existence de coûts induits par la présence d'enfants (coûts de transport, de garde,...). En fait, la césure la plus nette pour les tranches d'âge considérées concerne les femmes qui ont trois enfants ou davantage (tableau 1). En deçà, les différences ne sont pas significatives. Ce résultat contraste avec d'autres publications qui mettent en évidence une activité moindre pour les femmes, dès qu'elles sont mères de deux enfants ou plus. L'écart provient du fait que l'on s'intéresse ici à des femmes qui, pour la quasi-totalité, ont donné naissance à tous leurs enfants. Les publications usuelles concernent les femmes adultes de tous âges. Parmi celles-ci, l'ensemble des mères de deux enfants contient des femmes dont la descendance finale sera effectivement de deux enfants, mais aussi des femmes qui auront ultérieurement un troisième enfant.

TABLEAU 2
Nombre total d'enfants par femme et nombre moyen d'enfants vivant dans le logement

Âge	Nombre moyen d'enfants	Nombre moyen d'enfants dans le logement
Moins de 25 ans	0,53	0,51
De 25 à 29 ans	1,11	1,10
De 30 à 34 ans	1,90	1,85
De 35 à 39 ans	2,30	2,24
De 40 à 44 ans	2,38	2,12
De 45 à 49 ans	2,21	1,42
De 50 à 54 ans	2,43	0,86
55 ans et 56 ans	2,44	0,50
Ensemble	1,98	1,54

Champ : Femmes conjoint de salariés, âgées de 56 ans au moins.
Source : Panel européen 1994, Insee.

TABLEAU 3
Modèle probit simple sur le comportement d'activité

	Coefficient	Écart-type
Constante	0,86	0,06
Pas de diplôme	-0,39	0,08
Avoir un Bac technique	0,47	0,20
Avoir un diplôme supérieur au Bac	0,53	0,11
Avoir de 50 à 54 ans	-0,26	0,09
Avoir 55 ans ou 56 ans	-0,51	0,13
Le conjoint gagne plus de 19 800 F	-0,43	0,10
Avoir eu une mère active pendant l'enfance	0,28	0,07
Avoir trois enfants ou plus	-0,71	0,07
-2 Log-vraisemblance	1 960,21	

Champ : Femmes conjoint de salariés âgées de 35 à 56 ans.
Source : Panel européen 1994, Insee.

Or, l'examen des données longitudinales suggère que les choix d'activité sont davantage conditionnés par la descendance projetée que par la descendance à l'instant présent (LOLLIVIER [2000]). Le moindre taux d'activité des futures mères de trois enfants pèserait alors sur le taux apparent d'activité des femmes ayant deux enfants. Ceci expliquerait pourquoi, dans l'échantillon considéré dans l'étude, la seule différence porte sur les mères de trois enfants et plus.

Dans un sens, ceci simplifie la formulation du problème d'endogénéité, puisque l'interaction entre activité et fécondité peut être traitée au moyen de modèles portant sur deux variables dichotomiques : le choix d'activité et le fait d'avoir trois enfants ou davantage. La question de l'endogénéité de la fécondité peut ainsi être modélisée au moyen d'un modèle probit bivarié, sans qu'il soit nécessaire d'avoir recours à une représentation polytomique.

3.3 L'estimation

Pour mener à bien les tests d'endogénéité et l'estimation du modèle bivarié, on a conservé dans les ajustements les seules variables explicatives significatives[1]. Parmi celles-ci, deux ne jouent que sur la fécondité nombreuse (tableau 4) : le fait d'être né hors de France (positivement) et la profession du père de la femme pendant son enfance ; les filles d'indépendant ont moins fréquemment trois enfants ou plus et les filles d'agriculteurs plutôt plus souvent trois enfants ou plus (le coefficient est à peine significatif). *A contrario*, le revenu du conjoint n'influence que la décision d'activité, de même que l'âge de la femme (tableau 3).

Pour analyser l'endogénéité de la fécondité dans la décision d'activité, on a à la fois calculé la valeur du test et réalisé l'estimation du modèle bivarié par la méthode du maximum de vraisemblance. Ce dernier permet de réaliser le test du maximum de vraisemblance sur le paramètre d'interaction entre les termes d'erreur. Concrètement, il est préférable de se livrer à l'estimation du

TABLEAU 4

Modèle probit simple sur le fait d'avoir trois enfants ou plus

	Coefficient	Écart-type
Constante	-0,51	0,05
Pas de diplôme	0,62	0,08
Avoir un CEP	0,36	0,08
Être née hors de France	0,29	0,09
Père agriculteur pendant l'enfance	0,17	0,09
Père indépendant non agricole	-0,22	0,11
Avoir eu une mère active pendant l'enfance	-0,17	0,06
-2 Log-vraisemblance	2 310,67	

Champ : Femmes conjoint de salariés âgées de 35 à 56 ans.
Source : Panel européen 1994, Insee.

1. La prise en compte de toutes les variables dans les ajustements ne change en rien les résultats, mais alourdit sensiblement la présentation.

modèle bivarié en se livrant à un balayage portant sur les valeurs du paramètre ρ_2 et maximisant la log-vraisemblance partielle conditionnelle à la valeur de ρ_2. En effet, comme le soulignent des travaux portant sur des problèmes analogues, la log-vraisemblance n'est pas monotone et fait apparaître des extrema locaux. On procède donc de façon identique à celle préconisée pour estimer le modèle Tobit généralisé. En termes latents, on notera d'ailleurs que les deux modèles sont très proches, comme l'a remarqué HECKMAN [1978] : biais de sélectivité et d'endogénéité renvoient à une formulation commune.

Le test fondé sur la régression auxiliaire rejette très nettement l'hypothèse alternative d'endogénéité de la variable de fécondité dans le modèle d'activité : la valeur obtenue, de l'ordre de 0,6, est très inférieure à la valeur critique (tableau 5). Ce résultat est confirmé par l'estimation par la méthode du maximum de vraisemblance : celui-ci fournit une valeur optimale de ρ voisine de 0,5, mais à l'optimum, la log-vraisemblance ne diffère qu'au niveau de la deuxième décimale de la valeur correspondante sous l'hypothèse nulle (tableau 6). Dans l'estimation, la variable de fécondité apparaît donc clairement exogène.

TABLEAU 5
Régressions auxiliaires et tests d'endogénéité

	Modèle complet		Modèle sans la variable d'activité de la mère	
	Coefficient	Écart-type	Coefficient	Écart-type
Variables du modèle d'activité				
Constante	0,12	0,17	0,31	0,15
Pas de diplôme	0,08	0,13	0,21	0,12
Avoir un Bac technique	-0,01	0,21	-0,03	0,21
Avoir un diplôme supérieur au Bac	-0,01	0,12	-0,02	0,11
Avoir de 50 à 54 ans	0,00	0,09	0,01	0,08
Avoir 55 ans ou 56 ans	0,00	0,12	0,01	0,12
Le conjoint gagne plus de 19 800 F	-0,01	0,10	-0,03	0,10
Avoir eu une mère active pendant l'enfance	-0,02	0,07		
Avoir trois enfants ou plus	-0,34	0,45	-0,92	0,42
Variables du modèle de fécondité				
Constante	0,00	0,05	0,00	0,05
Pas de diplôme	0,00	0,08	0,00	0,08
Avoir un CEP	0,00	0,08	0,01	0,08
Être née hors de France	0,00	0,09	0,00	0,09
Père agriculteur pendant l'enfance	0,00	0,09	0,01	0,09
Père indépendant non agricole	0,00	0,11	0,01	0,11
Avoir eu une mère active pendant l'enfance	0,00	0,07	0,00	0,07
ρ	0,21	0,27	0,57	0,26
Somme des carrés de la variable expliquée	0,59		5,06	

Lecture : Les régressions consistent à ajuster un vecteur composé de 1 sur les valeurs prises par les dérivées de la vraisemblance en $\rho = 0$. La somme des carrés de la variable expliquée suit un $\chi^2(1)$, dont la valeur critique est de 3,84. Dans le premier modèle, comportant toutes les variables, le test rejette l'endogénéité. En revanche, dans le second modèle, où une variable a été omise, le test accepte l'endogénéité.
Champ : Femmes conjoint de salariés âgées de 35 à 56 ans.
Source : Panel européen 1994, Insee.

Afin de procéder à une variante, on s'est livré à l'exercice consistant à conserver le même modèle en retirant de la liste des variables explicatives de l'activité le fait que la mère de la personne était active. Il s'agit là d'une importante erreur de spécification. Cette fois, le test obtenu à partir de la régression auxiliaire prend une valeur de 5,1, supérieure à la valeur critique, et qui conduit à rejeter l'hypothèse nulle (tableau 5). À nouveau, le résultat

TABLEAU 6

Estimation du modèle avec endogénéité, par la méthode du maximum de vraisemblance

	Probit séparés		Modèle avec endogénéité		
	Coefficient	Écart-type	Coefficient	*Coefficient normalisé*	Écart-type
Variables du modèle d'activité					
Constante	0,86	0,06	1,14	*1,02*	0,07
Pas de diplôme	-0,39	0,08	-0,30	*-0,26*	0,09
Avoir un Bac technique	0,47	0,21	0,48	*0,43*	0,23
Avoir un diplôme supérieur au Bac	0,53	0,11	0,55	*0,49*	0,12
Avoir de 50 à 54 ans	-0,26	0,08	-0,28	*-0,25*	0,09
Avoir 55 ans ou 56 ans	-0,51	0,12	-0,54	*-0,49*	0,13
Le conjoint gagne plus de 19 800 F	-0,43	0,10	-0,49	*-0,43*	0,11
Avoir eu une mère active pendant l'enfance	0,28	0,07	0,28	*0,25*	0,07
Avoir trois enfants ou plus	-0,71	0,07	-1,34	*-1,20*	0,07
Variables du modèle de fécondité					
Constante	-0,51	0,05	-0,51		0,05
Pas de diplôme	0,62	0,08	0,63		0,08
Avoir un CEP	0,36	0,08	0,37		0,08
Être née hors de France	0,29	0,09	0,28		0,09
Père agriculteur pendant l'enfance	0,17	0,09	0,17		0,09
Père indépendant non agricole	-0,23	0,11	-0,22		0,11
Avoir eu une mère active pendant l'enfance	-0,17	0,07	-0,17		0,06
ρ_2	0		0.5		
– 2 Log-vraisemblance	4 270,88		4 270,84		

Lecture : Le tableau présente l'ajustement du modèle avec toutes les variables. Les colonnes « probit séparés » fournissent l'estimateur du maximum de vraisemblance lorsque $\rho_2 = 0$. Les colonnes « modèle avec endogénéité » fournissent l'estimation correspondante à la valeur de ρ_2 qui maximise la vraisemblance, 0,5 en l'occurrence (valeur déterminée par balayage sur ρ_2). La colonne « coefficient normalisé » fournit à titre indicatif la valeur du coefficient du modèle, après division par l'écart-type du terme d'erreur de la première équation ($\sqrt{1+\rho_2^2}$), afin de pouvoir comparer la valeur avec celle du modèle sans hétérogénéité. On notera qu'en dépit de coefficients parfois un peu différents (voire même significativement différents pour la variable de fécondité), l'écart entre le double des log-vraisemblances est plus petit que la valeur prise par un $\chi^2(1)$, ce qui conduit à rejeter l'endogénéité.

Champ : Femmes conjoint de salariés âgées de 35 à 56 ans.

Source : Panel européen 1994, Insee.

est confirmé par l'estimation par la méthode du maximum de vraisemblance, qui lui aussi rejette l'hypothèse nulle (tableau 7). Le maximum de vraisemblance est atteint pour une valeur de ρ_2 égale à 4 (correspondant à une corrélation entre les ε voisine de 0,7) et prend la valeur 2 141,38 à comparer à une valeur de 2 144,47 sous l'hypothèse nulle.

TABLEAU 7

Estimation du modèle avec endogénéité, avec omission d'une variable explicative

	Probit séparés		Modèle avec endogénéité		
	Coefficient	Écart-type	Coefficient	*Coefficient normalisé*	Écart-type
Variables du modèle d'activité					
Constante	0,98	0,05	4,78	*1,16*	0,14
Pas de diplôme	-0,41	0,08	-0,35	*-0,08*	0,24
Avoir un Bac technique	0,45	0,21	1,19	*0,29*	0,88
Avoir un diplôme supérieur au Bac	0,53	0,11	1,57	*0,38*	0,40
Avoir de 50 à 54 ans	-0,26	0,08	-0,81	*-0,20*	0,25
Avoir 55 ans ou 56 ans	-0,52	0,12	-1,35	*-0,33*	0,36
Le conjoint gagne plus de 19 800 F	-0,43	0,10	-1,67	*-0,41*	0,34
Avoir trois enfants ou plus	-0,72	0,07	-7,10	*-1,72*	0,26
Variables du modèle de fécondité					
Constante	-0,51	0,05	-0,48		0,05
Pas de diplôme	0,62	0,08	0,64		0,08
Avoir un CEP	0,36	0,08	0,34		0,07
Être née hors de France	0,29	0,09	0,23		0,08
Père agriculteur pendant l'enfance	0,17	0,09	0,15		0,09
Père indépendant non agricole	-0,23	0,11	-0,17		0,10
Avoir eu une mère active pendant l'enfance	-0,17	0,07	-0,24		0,06
ρ_2	0		4		
– 2 Log-vraisemblance	4 288,94		4 282,76		

Lecture : Le tableau présente l'ajustement du modèle après omission de la variable « Avoir eu une mère active pendant l'enfance ». La lecture est analogue à celle du tableau 6. On notera cette fois que l'on accepte l'endogénéité. L'omission d'une variable fait donc apparaître une corrélation entre les termes d'erreur des deux équations.
Champ : Femmes conjoint de salariés âgées de 35 à 56 ans.
Source : Panel européen 1994, Insee.

4 Conclusion

Plusieurs enseignements découlent de cette analyse. En premier lieu, la décision de fécondité apparaît exogène lorsque l'on cherche à évaluer l'influence des enfants sur l'activité, sauf à commettre de graves erreurs de spécifications. On notera que l'on n'a pas interprété ce résultat en termes de causalité, le modèle estimé ne dérivant pas d'un modèle structurel. Ce que montre l'estimation du modèle bivarié est que l'équation de participation au marché du travail comporte suffisamment de variables explicatives pour qu'aucune variable inobservée ne crée de corrélation entre la variable de fécondité et le terme d'erreur.

En second lieu, il faut noter la relative simplicité de mise en œuvre de ces techniques. Le test conduit à partir des régressions auxiliaires ne nécessite aucune programmation spécifique de la part de l'utilisateur. En outre, la méthode semble donner de bons résultats appliquée à ce type de problème, même si dans la littérature, certains auteurs sont assez réservés quant à son utilisation. La raison provient peut-être de ce que l'échantillon présente une taille suffisante, ou encore que le test est bien adapté à la nature du problème. Compte tenu de la simplicité de sa mise en œuvre, on serait plutôt tenté de recommander son utilisation sur données individuelles, plutôt que de supposer purement et simplement l'exogénéité de certains comportements, comme c'est souvent le cas dans les études utilisant des variables qualitatives.

Enfin, l'estimation du maximum de vraisemblance par la méthode décrite présente de nombreux avantages par rapport au calcul direct de la vraisemblance du modèle bivarié. Il est vrai que la fonction de répartition de la loi normale bivariée est disponible dans de nombreux logiciels (GAUSS, SAS), ce qui permet de calculer directement la vraisemblance. Ceci étant, la détermination du maximum de vraisemblance est beaucoup plus coûteuse en ressources que par la méthode préconisée ici. Pour mémoire, il suffit de quelques secondes à un micro-ordinateur standard pour réaliser les ajustements décrits dans l'article. En outre, la méthode décrite ici permet de calculer analytiquement la dérivée de la vraisemblance par rapport aux paramètres, et d'en déduire la matrice d'information de *Fisher*. Pour sa part, le recours à un calcul direct de la vraisemblance au moyen de la fonction de répartition bivariée impose le calcul numérique des dérivées afin de pouvoir réaliser des tests. De surcroît, la méthode peut s'appliquer à d'autres catégories de problèmes, faisant intervenir par exemple des modèles polytomiques, pour lesquels l'estimation directe n'est plus envisageable. ■

• Références bibliographiques

ABRAMOWITZ M., STEGUN I. (1964). – « Handbook of Mathematical Functions with Formula, Graphs, and Mathematical Tables », *National Bureau of Standard Applied Mathematics Series*, N° 55, Washington DC: US Government Printing Office.

BALTAGI B. (1999). – « Specification Tests in Panel Data Models Using Artificial Regressions », *Annales d'Économie et de Statistique*, Vol 55-56, pp. 277-297.

BOURGUIGNON F. (1986). – « Female Participation and Taxation in France », *Unemployment, Search and Labour Supply*, Blundell and Walker Eds, Cambridge University Press, pp. 243-266.

BUTLER J., MOFFITT R. (1982). – « A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model », *Econometrica*, Vol 50, N° 3, pp. 761-764.

DAGSVIK J., LAISNEY F., STROM S., OSTERVOLD J. (1988). – « Female Labour Supply and the Tax Benefit System in France », *Annales d'Économie et de Statistique*, N° 11, pp. 5-40.

MAC KINNON J. (1992). – « Model Specification Tests and Artificial Regressions », *Journal of Economic Literature*, Vol XXX, pp. 102-146.

DAVIDSON R, MAC KINNON J. (1990). – « Specification Tests Based on Artificial Regressions », *Journal of the American Statistical Association*, Vol 85, pp. 220-227.

GODFREY L., MCALEER M., MC KENZIE C. (1988). – « Variable Addition and Lagrange Multiplier Tests for Linear and Logarithmic Regression Models », *Review of Economic Studies*, N° 48, pp. 487-496.

HECKMAN J. (1974). – « Shadow Prices, Market Prices and Labor Supply », *Econometrica*, Vol 42, pp. 679-693.

HECKMAN J. (1978). – « Dummy Endogenous Variables in a Simultaneous Equation System », *Econometrica*, Vol 46, pp. 931-959.

HECKMAN J., SINGER B. (1984). – « A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data », *Econometrica*, Vol 52, pp. 271-320.

LEE L-F, MADDALA G., TROST R. (1980). – « Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity », *Econometrica*, Vol 48, pp. 491-503.

LOLLIVIER S. (2000). – « Activité féminine : une approche longitudinale », *Miméo*.

MROZ T. (1999). – « Discrete Factor Approximation in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on Continuous Outcome », *Journal of Econometrics*, Vol 92, pp. 233-274.

ROBIN J.M. (2000). – « Modèles structurels et variables explicatives endogènes », *Méthodologie Statistique*, N° 2002, INSEE.

WHITE H. (1980). – « A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity », *Econometrica*, Vol 48, pp. 817-838.

ANNEXE

Calcul de la valeur du gradient sous l'hypothèse nulle = 0

Afin d'alléger les formules, on notera :

$$A_1 = Y_1^* - X_1\beta_1$$

$$A_2 = \frac{X_2\beta_2 + \alpha Y_1 + \rho(Y_1^* - X_1\beta_1)}{\sqrt{1-\rho^2}}$$

La contribution d'un individu à la log-vraisemblance s'écrit alors :

$$l = \mathrm{Log}(\mathcal{L}) = \mathrm{Log}\left[\int_{D(Y_1)} Z_2(A_2)\varphi(A_1)dY_1^*\right]$$

En $\tilde{\theta} = (\tilde{\beta}_2, \tilde{\alpha}, \tilde{\beta}_1, 0)$, elle se réduit à :

$$\tilde{l} = \mathrm{Log}(\tilde{\mathcal{L}}) = \mathrm{Log}(Z_2(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)) + \mathrm{Log}(\mathrm{Z}_1(\mathrm{X}_1\tilde{\beta}_1))$$

c'est-à-dire la somme des log-vraisemblances de deux modèles probits séparés.

On note enfin :

$$DL(A_i) = \frac{\varphi(A_i)(Y_i - \Phi(A_i))}{\Phi(A_i)(1-\Phi(A_i))}$$

pour $i = 1,2$. L'expression DL correspond à la dérivée par rapport à la constante de la log-vraisemblance d'un modèle probit simple.

Par conséquent,

$$\frac{\partial l}{\partial \beta_2} = \frac{1}{\mathcal{L}}\int_{D(Y_1)} Z_2(A_2)\frac{\partial \mathrm{Log}(Z_2(A_2))}{\partial \beta_2}\varphi(A_1)dY_1^*$$

$$= \frac{X_2'}{\mathcal{L}}\int_{D(Y_1)} Z_2(A_2)DL(A_2)\varphi(A_1)dY_1^*$$

En $\tilde{\theta} = (\tilde{\beta}_2, \tilde{\alpha}, \tilde{\beta}_1, 0)$,

$$\frac{\partial l}{\partial \beta_2} = \frac{X_2'}{\tilde{\mathcal{L}}}Z_2(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)DL(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)\int_{D(Y_1)}\varphi(Y_1^* - X_1\tilde{\beta}_1)dY_1^*$$

$$= DL(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)X_2'$$

en simplifiant par $\tilde{\mathcal{L}} = Z_1(X_1\tilde{\beta})Z_2(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1)$ au numérateur et au dénominateur. On montre de la même façon que :

$$\frac{\partial l}{\partial \alpha} = DL(X_2\tilde{\beta}_2 + \tilde{\alpha}_1)Y_1$$

et

$$\frac{\partial l}{\partial \rho} = \frac{1}{\mathcal{L}} \int_{D(Y_1)} Z_2(A_2) \frac{\partial \text{Log}(Z_2(A_2))}{\partial \rho} \varphi(A_1) dY_1^*$$
$$= \frac{1}{\mathcal{L}} \int_{D(Y_1)} Z_2(A_2) DL(A_2) \frac{\partial A_2}{\partial \rho} \varphi(A_1) dY_1^*$$

avec,

$$\frac{\partial A_2}{\partial \rho} = \frac{A_1(1-\rho^2)^{1/2} + A_2\rho(1-\rho^2)^{-1/2}}{(1-\rho^2)}$$

En $\tilde{\theta} = (\tilde{\beta}_2, \tilde{\alpha}, \tilde{\beta}_1, 0)$, l'expression de la dérivée se simplifie notablement :

$$\frac{\partial l}{\partial \rho} = \frac{1}{\tilde{\mathcal{L}}} Z_2(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1) DL(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1) \int_{D(Y_1)} (Y_1^* - X_1\tilde{\beta}_1) \varphi(Y_1^* - X_1\tilde{\beta}_1) dY_1^*$$
$$= \frac{1}{\tilde{\mathcal{L}}} Z_2(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1) DL(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1) \left[-\varphi(Y_1^* - X_1\tilde{\beta}_1) \right]_{D(Y_1)}$$

On vérifie que :

$$\frac{1}{Z_1(X_1\tilde{\beta}_1)} \left[-\varphi(Y_1^* - X_1\tilde{\beta}_1) \right]_{D(Y_1)} = DL(X_1\tilde{\beta}_1)$$

Donc,

$$\frac{\partial l}{\partial \rho} = DL(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1) DL(X_1\tilde{\beta}_1)$$

Enfin,

$$\frac{\partial l}{\partial \beta_1} = \frac{1}{\mathcal{L}} \int_{D(Y_1)} Z_2(A_2) DL(A_2) \frac{\partial A_2}{\partial \beta_1} \varphi(A_1) dY_1^* + \frac{1}{\mathcal{L}} \int_{D(Y_1)} Z_2(A_2) \varphi'(A_1) \frac{\partial A_1}{\partial \beta_1} dY_1^*$$

En $\tilde{\theta} = (\tilde{\beta}_2, \tilde{\alpha}, \tilde{\beta}_1, 0)$, le premier terme de la somme s'annule. Il reste :

$$\frac{\partial l}{\partial \beta_1} = \frac{X_1'}{\tilde{\mathcal{L}}} Z_2(X_2\tilde{\beta}_2 + \tilde{\alpha}Y_1) \left[-\varphi(Y_1^* - X_1\tilde{\beta}_1) \right]_{D(Y_1)}$$

soit :

$$\frac{\partial l}{\partial \beta_1} = X_1' DL(X_1\tilde{\beta}_1)$$