Recueil du symposium 2002 de Statistique Canada
Modélisation des données d’enquête pour la recherche économique et sociale

INTÉGRATION DE LA THÉORIE DE LA RÉPONSE AUX ITEMS AUX MODÈLES PAR ÉQUATIONS STRUCTURELLES : COMPARAISON AVEC UNE RÉGRESSION FONDÉE SUR DES SCORES TRI

D.R. Thomas[1], I.R.R. Lu[1] et B.D. Zumbo[2]

RÉSUMÉ

Les auteurs passent en revue les problèmes d’utilisation de scores TRI de variables latentes, examinent le lien entre les modélisations respectives de régression de variables latentes TRI et MES dans le cas de données discontinues et comparent les estimations paramétriques de régression par les scores prévues TRI et les scores normalisées «de réponses justes» dans une régression MCO dont les estimations sont tirées de la double méthode TRI-MES. Les résultats de Monte Carlo indiquent que le cadre PAP est insensible à la taille d’échantillon comme on pouvait le prévoir, mais aboutit à une atténuation appréciable des estimations paramétriques de régression. Quant à la méthode TRI-MES, elle produit un moindre biais d’échantillon fini et, comme prévu, donne des estimations «convergentes» pour des tailles d’échantillon d’un ordre de grandeur approprié.

MOTS CLÉS : TRI; MES; modélisation à variables latentes; scores «de réponses justes»; statistiques personnelles; scores.

1. INTRODUCTION

Dans les sciences du comportement, les chercheurs étudient fréquemment des concepts qui ne se prêtent pas à une observation directe, qu’il s’agisse de la satisfaction au travail, du stress professionnel, des niveaux de dépression clinique, de l’aptitude à compter ou à lire chez les enfants ou de la compétence en gestion, pour ne citer que ces exemples. Pour mesurer ces concepts que nous appellerons «variables latentes», on obtient des données de réponse à un ensemble d’items à échelle dans les divers domaines en question. S’il y a lieu, on peut alors disposer de scores individuelles de variables latentes par diverses techniques comme celles de l’analyse factorielle, des échelles psychométriques classiques et de la théorie de la réponse aux items (TRI). Toutefois, l’intérêt en recherche se porte habituellement sur les relations entre variables latentes, et non pas sur des mesures de ces variables mêmes. On a conçu un éventail de techniques pour faciliter cette modélisation à variables latentes, et notamment la modélisation par équations structurelles (MES) qu’offrent aujourd’hui des programmes comme LISREL (Joreskog et Sorbom, 1996), EQS (Bentler, 1995), AMOS (Arbuckle, Wothke, 1999), Mplus (Muthen et Muthen, 2001) et LINCS (Schoenberg et Arminger, 1990). En joignant des modèles de mesure en analyse factorielle à des équations structurelles du type «régression», les méthodes MES permettent des estimations convergentes des paramètres d’un ensemble très général de modèles. Il existe aujourd’hui sur les méthodes MES une vaste documentation spécialisée qui recoupe les études consacrées aux modèles d’erreurs de mesure (Fuller, 1987) et qui, dans bien des cas, les développe. Il reste que, si on fait un usage croissant de la méthodologie MES surtout pour les données en mesure continue, maints chercheurs préfèrent une méthode plus simple où on obtient des scores de variables latentes et les intègre directement aux analyses de régression et aux analyses multivariées. Très répandue, cette méthode n’en présente pas moins de graves inconvénients. On sait bien que le recours direct à des scores de variables latentes en régression et dans d’autres analyses statistiques peut avoir pour résultat des estimations divergentes ou biaisées des paramètres du modèle et, là où les échelles de variables latentes sont fondées sur trop peu d’items, le biais peut prendre de sérieuses proportions. Un tel biais est probable lorsqu’on prévoit la valeur d’une variable latente dans un

[1]Sprott School of Business, Université Carleton, Ottawa, Ontario, Canada K1S 5B6 (adresse électronique : rthomas@sprott.carleton.ca).

[2]Université de Colombie-Britannique, pavillon Scarfe, 2125, Main Mall, Department of ECPS, Vancouver, Colombie-Britannique, Canada V6T 1Z4.

modèle à effets aléatoires. (À noter que, dans notre exposé, le terme *prévision* sert à distinguer la production de scores relatifs à des variables latentes et aléatoires de l'*estimation* de paramètres fixes de modèles.) En général, la distribution des valeurs prévues ne converge pas avec celle de la variable latente à mesure qu'augmente la taille d'échantillon sauf si le nombre d'item dans l'échelle est très élevé. Les analyses qui font appel aux valeurs prédites de variables latentes sont alors susceptibles d'être entachées d'un biais, et ce, sans égard à la taille d'échantillon (Little et Rubin, 1983; Louis, 1984). Il y a convergence seulement lorsque le nombre d'items dans l'échelle va à l'infini. C'est ce que nous appellerons le biais de nombre fini d'items dans notre exposé. On a beaucoup décrit les problèmes d'utilisation de scores de variables latentes tirées d'une analyse factorielle de données d'échelle continue dans une régression multiple. La grande conclusion de la recherche consacrée à la question est qu'une régression portant sur des scores en analyse factorielle risque d'engendrer un grave biais de nombre fini d'items (Tucker, 1971; Shevlin, Miles et Bunting, 1997), bien que cette déformation puisse être évitée dans certains cas (Skrondal et Laake, 2001) ou atténuée (Croon, 2002). Si on s'est beaucoup intéressé à tout ce qui est «échelles continues», on n'a guère prêté attention aux problèmes d'utilisation de données d'échelle binaire en analyse de régression.

Nous nous attacherons, par conséquent, aux questions d'utilisation de scores de variables latentes TRI en modélisation d'analyse. Nous examinerons non seulement les scores TRI, mais les scores normalisées «de réponses justes» (RJ) d'usage courant. Il s'agit là du nombre total d'items du test répondu juste pour une personne après une normalisation qui met la moyenne d'échantillon à zéro et l'écart-type d'échantillon à l'unité. À la section 2, nous résumerons la méthodologie de base TRI et décrirons les méthodes de prévision des variables latentes des modèles TRI. À la section 3, nous passerons en revue les études existantes sur l'utilisation de scores binaires TRI en régression MCO et dans d'autres analyses. Il sera brièvement question des méthodes de correction et de prévention du biais de nombre fini d'items. On a amplement scruté le rapport entre l'analyse factorielle et la méthodologie MES, mais on ne comprend pas aussi bien la relation du même ordre entre les cadres TRI et MES. À la section 4, nous cernerons donc ce lien et exposerons une méthode d'intégration des modèles de mesure TRI à une analyse de régression de variables latentes MES. Cette technique donne des estimations convergentes de tous les paramètres, c'est à dire qu'elle permet d'éviter le problème de biais de nombre fini d'items. Il sera toujours nécessaire cependant d'établir à quel point les estimations TRI-MES théoriquement convergentes donnent de bons résultats dans la pratique, c'est à dire de mesurer le biais d'échantillon fini. Enfin, nous présenterons les résultats d'une première étude de simulation où sont comparées les estimations paramétriques de régression par scores prévues TRI et normalisées NJ dans une régression MCO aux estimations de régression par la double technique TRI-MES.

2. PRÉVISION DES SCORES DE VARIABLES LATENTES PAR QUESTIONS D'ÉCHELLE DISCONTINUE

Dans cette section, nous traiterons principalement des techniques expressément conçues pour l'élaboration d'échelles de mesure par questions discontinues et, en particulier, de la famille générale des méthodes relevant de la théorie de la réponse aux items (TRI) avec les méthodes mises au point pour des données binaires et des questions ordinales à catégories multiples. Par souci de concision, nous nous attacherons aux seuls modèles binaires. À la section 3, nous aborderons la question de l'utilisation de scores binaires TRI dans des analyses de régression.

2.1 Modèle TRI binaire

Les méthodes relevant de la théorie de la réponse aux items (TRI) ont vu le jour dans les domaines des tests d'aptitudes en éducation et des tests psychométriques où la variable latente représente l'*aptitude* ou la *caractéristique* d'un sujet. Dans de tels tests, un sujet doit répondre à un certain nombre d'items B binaires pour nos besoins B et les réponses individuelles ainsi obtenues sont par la suite transformées en scores d'aptitude ou de caractéristique de la personne interrogée à l'aide d'un modèle TRI préalablement étalonné. Dans le contexte de la recherche par enquête, les items du test sont destinées à tous les enquêtés et les prévisions d'aptitudes individuelles (si on en obtient) et les paramètres du modèle TRI se calculent par le même ensemble de données d'enquête. Dans l'un et l'autre cas, la clé de l'application de méthodes TRI est un modèle qui relie les caractéristiques dégagées par un item précis du test et la juste valeur d'aptitude d'une personne à la probabilité que le sujet donne une juste réponse à cette question. Dans un modèle TRI binaire à courbe sigmoïde normale (Bock et Liebermann, 1970; Bock et Aitkin, 1981), la probabilité de juste réponse à la j^e question compte tenu de l'aptitude η d'un sujet devrait être

$$p_j(\eta) = \int_{-\infty}^{a'\eta + b} \phi(z)\, dz = \Phi(a'_j\eta + b_j), \qquad (1)$$

où η représente un vecteur d'aptitudes latentes du sujet, a_j et b_j les paramètres propres à chaque item du test et ϕ et Φ la densité normale et la fonction normale de distribution cumulative (f*dc*) respectivement. Dans l'hypothèse d'une indépendance locale, la probabilité conditionnelle d'observation d'un vecteur résultat x de *n* réponses au test est la suivante :

$$P(x\,|\,\eta) = \prod_{j=1}^{n} [p_j(\eta)]^{x_j} [1 - p_j(\eta)]^{1-x_j}, \qquad (2)$$

de sorte que la vraisemblance marginale d'une observation x devient

$$P(x) = \int P(x\,|\,\eta)\, g(\eta)\, d\eta, \qquad (3)$$

où $g(\eta)$ est la forme posée de la distribution marginale de population de η. Dans la forme la plus répandue du modèle TRI, on suppose que l'aptitude latente η est unidimensionnelle et que la *fdc* normale peut être approchée (ou remplacée) par la *fdc* logistique. C'est ce que l'on appelle le modèle logistique binaire à deux paramètres. D'autres versions, dont une version à trois paramètres du modèle logistique TRI, sont aussi d'usage courant.

On a mis plusieurs techniques au point pour l'estimation des paramètres individuels d'items, a_j et b_j, dont les plus connues et les plus répandues sont la technique du maximum de vraisemblance marginale (MVM) conçue par Bock et Liebermann (1970) et affinée par la suite par Bock et Aitkin (1981). Dans la version Bock-Aitkin, la marginalisation sur η est facilitée par le traitement de $g(\eta)$ comme discontinu, et on résout les équations de vraisemblance résultantes par l'algorithme EM. Des techniques plus récentes font intervenir des renseignements auxiliaires dans l'estimation (Mislevy, 1987); il y a aussi des méthodes MCCM d'estimation simultanée des paramètres individuels d'items et de prévision des aptitudes latentes (Patz et Junker, 1999).

2.2 Prévision des scores TRI

On a proposé diverses méthodes de prévision de la valeur de l'aptitude latente η d'un sujet compte tenu de ses résultats au test x. Au nombre de ces méthodes, on compte les «estimations» de maximum de vraisemblance (EMV) tirées du traitement de η comme paramètre fixe et d'une maximisation de la «vraisemblance» (2) où les questions localement indépendantes jouent le rôle d'observations indépendantes. L'estimateur pondéré de vraisemblance (EPV; Warm, 1989) est étroitement apparenté à l'EMRobert Philips presented "The theory and applications of the score function for determining the priority of follow up in the Annual Survey of Manufactures" at the SSC meetings in Halifax.
, mais présente de meilleures propriétés de biais lorsque η tend vers l'infini, le η en question étant le nombre d'items du test. Son biais conditionnel à η est $o(n^{-1})$ comparativement à $O(n^{-1})$ pour l'EMV (Lord, 1983). Des prédicteurs de variables latentes peuvent aussi s'obtenir à partir de la distribution postérieure de η par

$$P(\eta\,|\,U_i) = \frac{P(x_i\,|\,\eta)\, g(\eta)}{\int P(x_i\,|\,\eta)\, g(\eta)\, d\eta}. \qquad (4)$$

Deux prédicteurs de ce type sont l'*estimateur de maximum de vraisemblance a posteriori* (MAP) par maximisation de l'équation (4) pour chaque sujet, et l'*estimation de prévision a posteriori* (PAP), qui représente la moyenne de la distribution postérieure (4). Cette dernière valeur est habituellement déterminée par intégration de Gauss-Hermite de la distribution latente postérieure (voir Stroud et Sechrest, 1966). Il convient de noter qu'on procède habituellement en fixant les paramètres individuels d'items de l'estimation de vraisemblance et les paramètres de la distribution latente (s'il y en a) à leurs valeurs estimées. En d'autres termes, les prédicteurs fondés sur (4) seront du type bayesien empirique.

Kim et Nicewander (1993) ont examiné en détail le biais, les erreurs-types et la fiabilité de ces prédicteurs de variables latentes par les techniques de Monte Carlo. Ils ont aussi regardé les scores normalisés NJ. Ils en ont conclu que, si les cinq résultats étaient d'une fiabilité à peu près identique, tous étaient entachés d'un grand biais conditionnel pour $|\eta|>1$, η étant mesuré à une échelle $N(0,1)$. Dans ce cas, les pires étaient la cote NJ et l'EMV. Le biais conditionnel, la fiabilité et les erreurs-types étaient semblables pour l'EPV et les prédicteurs bayesiens MAP et PAP. L'EPV présentait un biais conditionnel un peu moindre, mais l'erreur-type était supérieure à celles des deux autres.

3. BIAIS DE NOMBRE FINI D'ITEMS DANS LA RÉGRESSION DE VARIABLES LATENTES

Dans cette section, nous examinerons le biais de nombre fini d'items d'estimations paramétriques que cause l'utilisation de scores de variables latentes TRI en analyse statistique standard. Nous décrirons brièvement aussi les méthodes de correction ou d'élimination de ce biais par absence de prévision de scores individuels de variables latentes. L'analyse de régression nous intéressera le plus à cet égard, bien que certaines des méthodes proposées d'élimination du biais en question soient applicables à d'autres analyses.

3.1 Utilisation directe de scores TRI

Les exemples d'utilisation directe de scores TRI en régression sont moins fréquents dans les études spécialisées que les exemples de recours en analyse factorielle aux scores tirées d'items d'échelle continue, ce dont on ne s'étonnera pas puisqu'on ne connaît pas aussi bien la modélisation TRI que l'analyse factorielle en dehors des domaines de la psychométrie et des tests d'aptitudes en éducation. Tous semblent s'accorder à dire dans les études spécialisées que l'utilisation directe de scores TRI n'est pas à recommander (voir, par exemple, Mislevy, Johnson et Muraki, 1992; Hoijtink et Boomsma, 1996), créant des biais d'estimation de paramètres et d'erreurs-types, bien que les évaluations empiriques de ces effets soient rares. Une étude qui livre des données empiriques a été réalisée par Adams, Wilson et Wu (1997), qui ont regardé un développement du modèle de Rasch qui porte sur un certain nombre de types de modèles, dont le modèle de traitement de crédit partiel (Masters, 1982) conçu pour le traitement de variables indicatrices catégoriques et ordonnées. Leur plan d'estimation tient compte de l'information auxiliaire, c'est à dire qu'ils ont voulu représenter la variable latente η comme fonction linéaire de données auxiliaires ***Y*** sur le sexe, la situation socio-économique, etc. Le modèle pour η était un modèle factoriel simplifié de régression sous la forme

$$\eta = \mathrm{Y}\,\beta + \zeta\,, \qquad (5)$$

où on posait un vecteur normal de perturbations ζ à moyenne nulle et à variance σ^2. La méthode de Adams et coll. (1997) a étroitement à voir avec celle qui était auparavant utilisée par Mislevy (1987). Elle diffère de celle de Bock et Aitkin (1981) en ce que plus de données antérieures sont posées pour η. Entre autres, Adams et coll. (1997) ont soumis à une comparaison trois estimateurs de β, à savoir (1) l'estimateur obtenu par une estimation simultanée *MV-EM* des paramètres individuels d'items et des coefficients de régression du modèle, (2) un estimateur à trois degrés où les paramètres individuels étaient d'abord estimés sans les renseignements auxiliaires, où ils servaient ensuite à produire des estimations PAP de la caractéristique latente η et où on déterminait enfin les paramètres β par régression MCO des scores PAP par rapport aux données auxiliaires ***Y***, et (3) un estimateur à deux degrés où les paramètres de régression étaient tirés d'un ensemble de valeurs plausibles de η (voir la section 3.3 plus loin). Leurs résultats de simulation démontrent que les estimations MCO fondées sur les scores PAP se trouvaient à sous-estimer tant R^2 que l'ordre de grandeur des coefficients de régression, alors que les estimations paramétriques issues d'une estimation simultanée (méthode 1) et d'un traitement par valeurs plausibles (méthode 3) étaient proches des valeurs vraies du modèle.

3.2 Utilisation de scores TRI en correction de biais

Dans le cas des scores de variables latentes fondées sur des variables indicatrices en mesure continue, Croon (2002) a fait valoir que, dans certaines situations, une opération en deux étapes pour une régression de variables latentes ou d'autres analyses pourrait se révéler avantageuse. Ainsi, il peut être plus facile de chercher des modèles de mesure appropriés et la juste forme fonctionnelle d'un modèle de régression à variables latentes si les modèles de mesure sont estimés séparément à partir du modèle de variables latentes. On peut user d'arguments semblables pour le cas de variables indicatrices discontinues où les scores TRI peuvent être substituées aux variables latentes. Si on doit employer la méthode en deux étapes de régression de variables latentes dans le cas discontinu, les méthodes d'intégration de scores TRI en correction de biais deviennent importantes.

Les résultats de Kim et Nicewander (1993) dont nous avons parlé indiquent que les prédicteurs TRI les moins entachés d'un biais sont les estimations de mode postérieur (MAP) et de moyenne postérieure (PAP) et l'estimation pondérée de vraisemblance (EPV). Hoijtink et Boomsma (1996) ont étudié tant empiriquement que théoriquement le biais des prédicteurs MAP et EPV et élaboré des expressions asymptotiques, justes à l'ordre $O(1/n)$, pour la moyenne conditionnelle et la variance de la variable latente η, la covariance entre η et une co-variable mesurée $\boldsymbol{y}$. Pour le prédicteur EPV, ces expressions peuvent d'emblée être évaluées et utilisées dans toutes sortes d'analyses, dont l'analyse de variance et la régression multiple. D'après une simulation en analyse de variance, Hoijtink et Boomsma (1996) ont conclu que la méthode de correction de biais asymptotique était nécessaire, mais qu'il fallait toujours au moins 15 éléments binaires dans le modèle TRI pour garder à un niveau acceptable les estimations de moyenne et d'erreur en analyse de variance.

3.3 Valeurs plausibles

La technique des *valeurs plausibles* diffère des techniques que nous avons décrites en ce que des ensembles multiples de scores sont fournis pour chaque variable latente, ce que l'on doit se garder d'interpréter comme des prédicteurs individuels. En fait, les valeurs plausibles représentent les imputations multiples (Rubin, 1987) tirées de la distribution prévisionnelle de la variable latente, celle-ci étant traitée dans ce cas comme données manquantes. C'est une technique conçue pour des enquêtes comme l'enquête bisannuelle américaine «National Assessment of Education Progress» (NAEP), où on demande aux élèves de répondre individuellement à aussi peu que huit questions d'un test de sorte que les prévisions par estimation de vraisemblance ou distribution postérieure n'offrent nullement une précision suffisante. La distribution prévisionnelle prend en compte la variabilité tenant à lutilisation d'un nombre fini d'items d'échelle. Ainsi, les estimateurs de valeurs plausibles, qui consistent en espérances estimées à l'égard de la distribution prévisionnelle, seront exempts de tout biais de nombre fini d'items. On trouvera un examen détaillé de la méthodologie des valeurs plausibles de la NAEP dans Mislevy, Johnson et Muraki (1992).

3.4 Estimation simultanée de paramètres individuels d'items et de paramètres de régression

Les travaux de Adams et coll. (1997) pour des données ordinales de traitement de crédit partiel et les travaux antérieurs de Mislevy (1987) pour des variables indicatrices binaires offrent des méthodes d'estimation simultanée (et convergente) de paramètres individuels d'items TRI et de paramètres de régression où on contourne le besoin de prévoir des scores de variables latentes et évite donc les biais de nombre fini d'items. Dans le contexte d'un modèle de Rasch, Zwinderman (1991) a procédé à une modélisation semblable d'estimation simultanée des paramètres de régression. Ces méthodes visent à accroître la précision de l'estimation en intégrant des données auxiliaires dans l'équation (5), qui figure une variable latente de réponse, mais aussi des variables prédicteurs mesurées sans erreur. Elles ne nous donnent donc pas une méthode générale d'analyse de modèles de régression à variables latentes. Il n'en existe pas moins des méthodes générales d'estimation paramétrique convergente de systèmes d'équations linéaires à variables latentes et à variables indicatrices discontinues. Elles peuvent servir à une estimation simultanée des paramètres individuels d'items TRI et des paramètres de modèles de régression à variables latentes, comme nous allons le décrire.

4. ESTIMATION SIMULTANÉE DE PARAMÈTRES TRI ET DE PARAMÈTRES DE RÉGRESSION DE VARIABLES LATENTES

Depuis le milieu de la décennie 1970, de nombreux auteurs ont consacré de vastes travaux (Christopherson, 1975; Muthen, 1983, 1984, etc.) au développement des techniques MES de traitement de variables indicatrices discontinues. La méthodologie actuellement appliquée dans des programmes comme Mplus (Muthen et Muthen, 2001) permet l'analyse de mélanges de variables discontinues et continues et comporte des techniques robustes d'estimation des erreurs-types et des statistiques d'ajustement qui ne s'appuient pas sur des indicateurs normaux multidimensionnels. Cette méthodologie est fort générale et joint la modélisation TRI à l'estimation simultanée tant des paramètres individuels d'items TRI que des paramètres de modèles de régression à variables latentes. Le lien à établir entre la modélisation TRI et la modélisation générale MES ne semble pas aussi connu que le lien correspondant entre les modèles de mesure en analyse factorielle et la modélisation MES dans le cas des variables continues, bien que la correspondance TRI ait très bien été décrite par Takane et de Leeuw (1987). Muthen (1988) a relié la modélisation TRI à des variables externes comme des indicateurs groupés, ainsi qu'à des variables continues secondaires. La correspondance TRI a été approfondie par la suite par Muthen, Kao et Burstein (1991).

4.1 Modèles de régression à variables latentes MES avec indicateurs binaires

Le modèle par équations structurelles MES se présente sous la forme

$$\eta = \Gamma\xi + \zeta \qquad (6)$$

où η est un vecteur de variables latentes de réponse, ξ un vecteur de variables explicatives latentes, Γ une matrice de coefficients de régression et ζ un vecteur de perturbations indépendantes de ξ avec la matrice des covariances Ψ. Pour une spécification complète du système (6), qui est un cas d'espèce du modèle général fourni par le programme Mplus (Muthen et Muthen, 2001), il faut une spécification de modèles de mesure pour η et ξ, à savoir

$$x^* = \Lambda_x \xi + \delta \qquad \text{et} \qquad y^* = \Lambda_y \eta + \varepsilon \quad , \qquad (7)$$

où x^* et y^* sont eux-mêmes des variables inobservables consistant en vecteurs d'items (ou encore de variables indicatrices) qui sont respectivement de dimension p et q et où les * charges + Λ_x et Λ_y sont formées de matrices de paramètres fixes. Les vecteurs δ et ε contiennent des perturbations aléatoires indépendantes des variables latentes et les unes des autres. Pour le simple modèle de population dont il sera question ici, la moyenne de x^* et y^* peut être fixée à zéro. Comme x^* et y^* sont inobservables, leur variance est arbitraire et habituellement fixée à l'unité pour plus de commodité. On modélise ensuite les variables indicatrices discontinues qui sont observables x et y par

$$x_j = 1 \ \text{ si } \ x^*_j \geq \tau_j \ , \quad x_j = 0 \ \text{ dans les autres cas}, \quad j = 1, \bullet\bullet\bullet, p,$$

$$y_j = 1 \ \text{ si } \ y^*_j \geq \tau_j \ , \quad y_j = 0 \ \text{ dans les autres cas}, \quad j = p+1, \bullet\bullet\bullet, p+q, \qquad (8)$$

où les τ représentent les seuils à estimer à partir des données et où p et q sont les nombres respectifs de variables indicatrices ou d'items dans les modèles de mesure de η et ξ.

L'estimation de maximum de vraisemblance des paramètres du modèle ci-dessus se révèle difficile, car il s'agit d'intégrer $p+q$ variables normales multivariées en corrélation sur l'espace dimensionnel $p+q$ délimité par les valeurs seuils. Voilà pourquoi nous procédons à une estimation par les moindres carrés généralisés à information limitée selon les probabilités marginales simples et doubles des variables indicatrices discontinues (Christopherson, 1975; Muthen, 1984). La convergence des estimations paramétriques a été établie.

4.2 Lien entre les modélisations TRI et MES

Takane et de Leeuw (1987) ont démontré qu'un modèle de mesure discontinue de la forme donnée à l'équation (8) équivaut formellement au modèle TRI en courbe sigmoïde normale donné à l'équation (1). Ils ont montré en particulier que, pour un modèle de mesure discontinue de la forme $y^* = \Lambda \eta + \varepsilon$, avec $var(\varepsilon) = \Theta$, les paramètres individuels d'items a_j et b_j de l'équation (1) peuvent s'exprimer dans les termes des paramètres du modèle de mesure discontinue par

$$a_j = \lambda_j / \theta_j^{1/2} \quad \text{et} \quad b_j = -\tau_j / \theta_j^{1/2}, \tag{9}$$

où λ_j est la j^e ligne de Λ et θ_j le j^e élément diagonal de Θ. Pour des caractéristiques latentes unidimensionnelles TRI, a_j sera scalaire. On peut donc considérer l'estimation des paramètres de la MES discontinue définie par les équations (6), (7) et (8) comme une estimation simultanée et convergente des paramètres individuels d'items TRI et des paramètres d'un modèle de régression à variables latentes. En d'autres termes, le modèle TRI est automatiquement intégré au modèle par équations structurelles. Dans tous les cas où les paramètres TRI a_j et b_j sont connus, il est possible de calculer les paramètres correspondants de modèle de mesure λ_j et τ_j, et de les garder fixes pendant l'estimation MES. C'est ce que nous appellerons la modélisation TRI-MES à paramètres fixes pour la distinguer d'une modélisation d'estimation simultanée tant des paramètres TRI que des paramètres d'équations structurelles. La TRI-MES à paramètres fixes livrera là encore des estimations convergentes des paramètres des équations structurelles.

5. ÉTUDE DE MONTE CARLO

L'étude de Monte Carlo comporte deux volets. En première étape, nous avons exploité les capacités de simulation du programme Mplus (Muthen et Muthen, 2001) pour examiner le biais d'échantillon fini dont est entachée l'estimation du paramètre Γ de régression de variables latentes par estimation simultanée et modélisation TRI-MES à paramètres fixes selon la description qu'en donne la section 4.2. En seconde étape, nous avons procédé à la prévision des scores PAP et NJ tant des variables latentes de réponse que des variables latentes explicatives à partir d'observations binaires simulées. Nous avons ensuite effectué une régression du paramètre de régression latente Γ par les moindres carrés ordinaires (MCO). Ce volet de l'étude vise à cerner le biais de nombre fini d'items attribuable à l'utilisation de scores prévues, et ce, en fonction du nombre d'items du test.

5.1 Conception de l'étude de Monte Carlo

Pour cette étude, nous avons choisi un modèle simple de régression par équations structurelles avec une variable latente de réponse η et une variable latente explicative ξ. Il s'agit de l'équation (6) où Γ est égal à un paramètre scalaire γ et fixé à la valeur $\sqrt{2}\,/\,2$. Les variances de η et de ξ fixées à l'unité donnent un coefficient de détermination de 0,5 pour le modèle par équations structurelles. Les modèles de mesure des variables latentes de réponse et explicatives sont $y_j^* = \lambda \eta_j + \varepsilon_j$ et $x_j^* = \lambda x_j + \delta_j$, $j = 1, \ldots, n$, où n est le nombre d'items de chaque échelle. Les paramètres de charge sont identiques. Enfin, les variances individuelles y_j^* et x_j^* sont fixées à l'unité.

La démarche de production de données de réponse aux items binaires aux fins de la simulation a comporté trois étapes. Premièrement, nous avons calculé les moyennes et les corrélations tétrachoriques de y_j^* et x_j^* pour une correspondance avec les modèles d'équations structurelles et de mesure. Deuxièmement, nous avons produit des données de réponse aux items normales multivariées à l'aide de ces moyennes et de ces corrélations tétrachoriques. Troisièmement, nous avons dichotomisé les données multivariées de réponse aux items en données binaires de réponse en nous reportant aux seuils que décrit la section 4.1. Ces valeurs seuils ont au départ été tirées d'une distribution normale type qui a été tronquée pour une fourchette de valeurs -1,5-+1,5, puis fixée pour la durée de l’étude. Dans ce premier volet, les deux dernières étapes décrites ont comporté une utilisation des capacités de simulation du programme Mplus.

Nous avons varié plusieurs caractéristiques à ce stade, à savoir a) les coefficients de détermination des modèles de mesure, c’est à dire *dc* = 0,3 et 0,6, b) le nombre d’items du test simulées, c’est à dire *n* = 10, 20 et 30, et c) les tailles d’échantillon, c’est à dire *N*=300, 500, 800 et 2 000. Pour chaque condition expérimentale, les résultats finals étaient fondés sur la moyenne de 1 000 itérations indépendantes.

Dans le second volet de l’étude, nous avons aussi produit des indicateurs binaires pour les variables latentes tant de réponse qu’explicatives. Nous avons ensuite calculé des valeurs de prévision PAP pour chacun des *N* cas d’une condition expérimentale donnée d’après le modèle TRI en courbe sigmoïde normale correspondant aux modèles de mesure utilisés en première étape (voir la section 4.2). Ainsi, les paramètres individuels d’items TRI ont été considérés comme connus et fixés aux fins de l’étude. Toutefois, les travaux de Kim et Nicewander (1993) semblent indiquer que les résultats obtenus par paramètres fixes ne différeront pas outre mesure de résultats par paramètres estimés pour les tailles d’échantillon qui sont ici les nôtres. Nous avons calculé les scores PAP par une intégration de Gauss-Hermite en 24 points de la distribution latente postérieure. Nous avons ensuite procédé par régression MCO par les scores PAP pour estimer le paramètre de régression de valeurs latentes γ. Les résultats sommaires étaient fondés sur 1 000 simulations indépendantes. Nous avons aussi repris cette étape de l’expérience en nous servant de scores NJ normalisées.

Dans les tableaux qui suivent, nous comparons les résultats théoriquement convergents tirés de la méthodologie TRI-MES aux résultats d’une régression de scores de variables latentes.

5.2 Résultats de l’étude de Monte Carlo

Le tableau 1 indique que, avec 10 questions dans chaque modèle de mesure, la méthode PAP est insensible à la taille d’échantillon comme on pouvait le prévoir, mais aboutit à une atténuation appréciable des estimations paramétriques de régression et des coefficients de détermination. Par ailleurs, la modélisation simultanée TRI-MES produit un moindre biais d’échantillon fini même pour les tailles inférieures d’échantillon et, comme prévu, elle livre des estimations * convergentes + de régression pour des tailles d’échantillon d’un ordre de grandeur approprié.

Le tableau 2 montre l’effet de la longueur du test sur le biais de nombre fini d’items dans les estimations de régression pour une taille d’échantillon 800. Comme on pouvait le prévoir, le biais paramétrique d’une régression PAP diminue à mesure qu’augmente le nombre d’items. Le tableau 2 indique en outre que plus s’élève le coefficient de détermination du modèle de mesure, plus rétrécit le biais de nombre fini d’items dans la régression PAP. Dans une estimation simultanée TRI-MES, ce biais était faible pour toutes les longueurs de test et se situait à un maximum approximatif de 2 % pour un test de 10 questions.

Tableau 1. Biais des estimations paramétriques de régression

Nombre d’items	Estimation de régression PAP			Estimation simultanée TRI-MES		
10-10		biais de en pourcentage	biais de en pourcentage		biais de en pourcentage	biais de en pourcentage
CD=0,3						
N=300	0,466 (0,002)	-34,1	-55,7	0,764 (0,01)	+8,06	+16,1
N=500	0,466 (0,001)	-34,1	-56,0	0,733 (0,007)	+3,68	+7,36
N=800	0,467 (0,001)	-33,9	-56,0	0,722 (0,005)	+2,12	+4,24
N=2 000	0,468 (0,001)	-33,8	-56,0	0,711 (0,003)	+0,57	+1,14
CD=0,6						
N=300	0,590 (0,001)	-16,6	-30,0	0,751 (0,006)	+6,22	+12,4
N=500	0,590 (0,001)	-16,6	-30,2	0,734 (0,004)	+3,82	+7,64
N=800	0,590 (0,001)	-16,6	-30,3	0,720 (0,003)	+1,84	+3,68
N=2 000	0,591 (0,001)	-16,4	-30,3	0,709 (0,002)	+0,28	+0,56

Note : N désigne la taille d’échantillon et CD, le coefficient de détermination de l’un et l’autre des modèles de mesure. Les erreurs-types de γ figurent entre parenthèses.

Tableau 2. Incidence de la longueur du test sur les estimations paramétriques de régression

N=800	Estimation de régression PAP			Estimation simultanée TRI-MES		
Nombre d'items		biais de en pourcentage	biais de en pourcentage		biais de en pourcentage	biais de en pourcentage
CD=0,3 10-10 20-20 30-30	 0,467 (0,001) 0,562 (0,001) 0,604 (0,001)	 -33,9 -20,5 -14,6	 -56,0 -36,6 -27,0	 0,722 (0,005) 0,711 (0,003) 0,714 (0,003)	 +2,12 +0,57 +0,99	 +4,24 +1,13 +1,98
CD=0,6 10-10 20-20 30-30	 0,590 (0,001) 0,640 (0,001) 0,660 (0,001)	 -16,6 -9,48 -6,65	 -30,3 -18,0 -13,0	 0,720 (0,003) 0,715 (0,003) 0,716 (0,003)	 +1,84 +1,13 +0,003	 +3,68 +2,26 +0,005

Le tableau 3 compare l'estimation simultanée à l'estimation TRI-MES à paramètres fixes. Il semblerait que la seconde crée un moindre biais d'échantillon fini que la première pour de moindres tailles d'échantillon. Même avec 10 questions et une taille d'échantillon de 300 seulement, le biais d'échantillon fini est inférieur à 5 % si les paramètres individuels d'items sont fixés à leurs valeurs connues.

Tableau 3. Comparaison des biais d'échantillon fini d'une estimation simultanée et d'une estimation TRI-MES à paramètres fixes

Nombre d'items : 20-20	Estimation TRI-MES à paramètres fixes			Estimation simultanée TRI-MES		
		biais de en pourcentage	biais de en pourcentage		biais de en pourcentage	biais de en pourcentage
CD=0,3 N=300 N=500 N=800 N=2 000	 0,739 (0,004) 0,724 (0,003) 0,717 (0,003) 0,710 (0,002)	 +4,53 +2,40 +1,41 +0,42	 +9,06 +4,80 +2,82 +0,84	 0,764 (0,01) 0,733 (0,007) 0,722 (0,005) 0,711 (0,003)	 +8,06 +3,68 +2,12 +0,57	 +16,1 +7,36 +4,24 +1,14
CD=0,6 N=300 N=500 N=800 N=2 000	 0,738 (0,002) 0,731 (0,002) 0,723 (0,002) 0,712 (0,001)	 +4,38 +3,39 +2,26 +0,71	 +8,76 +6,78 +4,52 +1,42	 0,751 (0,006) 0,734 (0,004) 0,720 (0,005) 0,709 (0,002)	 +6,22 +3,82 +1,84 +0,28	 +12,4 +7,64 +3,68 +0,56

Le tableau 4 indique que le biais de nombre fini d'items est à peu près le même pour les estimations normalisées NJ et les estimations de régression PAP. Une étude secondaire a démontré que les scores TRI et NJ étaient en corrélation presque parfaite.

Tableau 4. Comparaison des biais de nombre fini d'items des scores PAP et des scores normalisées NJ

CD	N	Nombre d'items : 10-10		Nombre d'items : 20-20	
		Scores PAP	Scores normalisées NJ	Scores PAP	Scores normalisées NJ
0,3	300 500 800	0,466 (0,002) 0,466 (0,001) 0,467 (0,001)	0,468 (0,001) 0,467 (0,001) 0,467 (0,001)	0,562 (0,002) 0,562 (0,001) 0,562 (0,001)	0,562 (0,001) 0,561 (0,001) 0,562 (0,001)
6	300 500 800	0,590 (0,001) 0,590 (0,001) 0,590 (0,001)	0,589 (0,001) 0,589 (0,001) 0,588 (0,001)	0,641 (0,001) 0,642 (0,001) 0,640 (0,001)	0,638 (0,001) 0,639 (0,001) 0,638 (0,001)

6. CONCLUSION

En général, la distribution des valeurs prévues ne converge pas vers la distribution de la variable latente à mesure qu'augmente la taille d'échantillon sauf là où le nombre d'items du test est suffisamment élevé. Les analyses fondées sur les valeurs prévues de variables latentes sont susceptibles d'être entachées d'un biais de nombre fini d'items, et il faut donc utiliser les scores de prévision de variables latentes avec circonspection.

Notre étude préliminaire de Monte Carlo indique que, pour des questions binaires, l'estimation simultanée et l'estimation à paramètres fixes TRI-MES produisent des estimations paramétriques de régression convergentes pour des tailles d'échantillon d'un ordre de grandeur approprié. Par ailleurs, si des scores TRI et des scores normalisées NJ servent à la régression, on a besoin d'un grand nombre d'items du test (n>30) pour ramener à des proportions acceptables le biais de nombre fini d'items dans les coefficients de régression. Ainsi, avec des échantillons assez grands, la modélisation TRI-MES peut être une solution de rechange à l'analyse et à l'estimation directes par scores TRI, ainsi qu'à l'application des méthodes de valeurs plausibles.

Comme ces dernières méthodes, la modélisation TRI-MES contourne la prévision de scores individuelles de variables latentes, mais à la différence de ces méthodes, elle n'exige ni un grand ensemble de variables de conditionnement, ni un gros effort de traitement de données de la part de l'organisme qui fournit les données. Ainsi, les méthodes TRI-MES peuvent être une autre façon pour des organismes comme Statistique Canada de rendre faisable une analyse convergente de régression de variables latentes pour les utilisateurs de leurs données. Les besoins en tailles d'échantillon d'une estimation simultanée TRI-MES peuvent être restrictifs dans certains contextes, mais le recours aux items préalablement étalonnés compensera cet effet dans une certaine mesure.

Une autre constatation de notre étude est que les estimations de régression par scores PAP et scores normalisées NJ sont hautement comparables, quelles que soient la longueur du test et la précision du modèle de mesure. Il faudra consacrer un complément de recherche à cette question.

RÉFÉRENCES

Adams, R. J., M. Wilson, et M. Wu. (1997), "Multilevel Item Response Models: An Approach to Errors in Variables Regression", *Journal of Educational and Behavioral Statistics*, 22(1), pp. 47-76.

Arbuckle, J. L. et W. Wothke (1999), *AMOS 4.0 User's Guide*, SPSS.

Bentler, P. M. (1995), *EQS Structural Equations Program Manual*, Encino, Calif.:Multivariate Software.

Bock, R. D. et M. Lieberman (1970), "Fitting a Response Model for N Dichotomously Scored Items", *Psychometrika*, 35(2), pp. 179-197.

Bock, R. D. et M. Aitkin (1981), "Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm", *Psychometrika*, 46(4), pp. 443-459.

Christoffersson, A. (1975), "Factor Analysis of Dichotomized Variables", *Psychometrika*, 40(1), pp. 5-32.

Croon, M. (2002), "Using Predicted Latent Scores in General Latent Structure Models", in G. A. Marcoulides and I. Moustaki (eds.) *Latent Variable and Latent Structure Models*, Mahwah, NJ: Lawrence Erlbaum Associates, pp. 195-223.

Fuller, W. A. (1987), *Measurement Error Models*, New York: Wiley.

Hoijtink, H. et A. Boomsma (1996), “Statistical Inference Based on Latent Ability Estimates”, *Psychometrika*, 61(2), pp. 313-330.

Jöreskog, K. G. et D. Sörbom (1996), *LISREL 8 User’s Reference Guide*, Chicago: Scientific Software International.

Kim, J. K. et W. A. Nicewander (1993), “Ability Estimation for Conventional Tests.” *Psychometrika*, 58(4), pp. 587-599.

Little, R. J. A. et D. B. Rubin (1983), “On Jointly Estimating Parameters and Missing Data by Maximizing the Complete Data Likelihood”, *American Statistician*, 37, pp. 218-220.

Lord, F. M. (1983), “Unbiased Estimators of Ability Parameters, of their Variance, and of their Parallel-forms Reliability”, *Psychometrika*, 48(2), pp. 233-245.

Louis, T. A. (1984), “Estimating a Population of Parameter Values Using Bayes and Empirical Bayes Methods”, *Journal of the American Statistical Association*, 79(386), pp. 393-398.

Masters, G. N. (1982), “A Rasch Model for Partial Credit Scoring”, *Psychometrika*, 47(2), pp. 149-174.

Mislevy, R. J. (1987), “Exploiting Auxiliary Information about Examinees in the Estimation of Item Parameters”, *Applied Psychological Measurement*, 11(1), pp.81-91.

Mislevy, R. J., E. G. Johnson, et E. Muraki (1992), “Scaling Procedures in Naep.” *Journal of Educational Statistics* 17(2), pp. 131-154.

Muthen, B. (1983), “Latent Variable Structural Equation Modeling with Categorical Data”, *Journal of Econometrics*, 22, pp. 43-65.

Muthen, B. (1984), “A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators”, *Psychometrika*, 49(1), pp. 115-132.

Muthen, B. (1988), “Some Uses of Structural Equation Modeling in Validity Studies: Extending IRT to External Variables”, in H. Wainer and H. I. Braun (eds.) *Test Validity*, Princeton, NJ: Educational Testing Service, pp. 213-238.

Muthen, B., C. F. Kao, et L. Burstein (1991), “Instructionally Sensitive Psychometrics: Application of a New IRT-Based Detection Technique to Mathematics Achievement Test Items”, *Journal of Educational Measurement*, 28(1), pp. 1-22.

Muthen, L. K. et B. O. Muthen (2001), *Mplus User’s Guide*, Los Angeles, CA: Muthen & Muthen.

Patz, R. J. et B. W. Junker (1999), “A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response models”, *Journal of Educational and Behavioral Statistics*, 24(2), pp. 146-178.

Rubin, D. B. (1987), *Multiple Imputation For Non-Response in Surveys*, New York: Wiley.

Schoenberg, R. et G. Arminger (1990). *LINCS2.0*. Kent. WA, RJS Software.

Shevlin, M., J. N. V. Miles, et B.P. Bunting (1997), “Summated Rating Scales: A Monte Carlo Investigation of the Effects of Reliability and Collinearity in Regression Models”, *Personality and Individual Differences*, 23(4), pp. 665-676.

Skrondal, A. et P. Laake (2001), “Regression among Factor Scores”, *Psychometrika*, 66(4), pp. 563-576.

Stroud, A. H. et D. Secrest (1966), *Gaussian Quadrature Formulas*, Englewood Cliffs, N.J.: Prentice-Hall.

Takane, Y. et J. Deleeuw (1987), “On the Relationship between Item Response Theory and Factor Analysis of Discretized Variables”, *Psychometrika*, 52(3), pp. 393-408.

Tucker, L. (1971), “Relations of Factor Score Estimates to their Use”, *Psychometrika*, 36(4), pp. 427-436.

Warm, T. A. (1989), “Weighted Likelihood Estimation of Ability in Item Response Theory”, *Psychometrika*, 54(3), pp. 427-450.

Zwinderman, A. H. (1991), “A Generalized Rasch Model for Manifest Predictors”, *Psychometrika*, 56(4), pp. 589-600.