« Un modèle des flux interrégionaux de marchandises au Canada »

Yvon Bigras et Sang Nguyen

L'Actualité économique, vol. 63, n° 1, 1987, p. 26-42.

DOI: 10.7202/601399ar

L'Actualité Économique. Revue d'analyse économique, vol. 63, n° 1, mars 1987

UN MODÈLE DES FLUX INTERRÉGIONAUX DE MARCHANDISES AU CANADA

Yvon BIGRAS,
Département d'administration et d'économique
Université du Québec à Trois-Rivières
et
Sang NGUYEN,
Centre de recherche sur les transports
Université de Montréal

Dans cet article, nous présentons un modèle des flux de marchandises entre huit régions du Canada, et pour l'ensemble des biens qui ont été regroupés en 64 catégories. Le modèle fonctionne en deux étapes. Dans un premier temps, les flux observés sont régressés sur certaines variables socio-économiques, dont le coût de transport. On obtient alors des flux « a priori », qui peuvent être modifiés lorsque les variables les expliquant sont elles-mêmes modifiées. Or, ces flux « a priori » ne respectent pas nécessairement la structure industrielle de chaque région. Pour les corriger, on résout un programme mathématique dont la fonction objectif est basée sur la théorie de l'information. On cherche alors les flux qui sont les plus proches possibles des flux « a priori », mais qui respectent également la structure industrielle de chaque région. Cette structure est représentée par des contraintes comptables input-output régionales. Le modèle peut être vu comme un modèle input-output interrégional, où les coefficients interrégionaux sont sensibles à des variations des coûts de transport. La formulation du modèle est également beaucoup plus souple et elle permet de prendre en compte d'autres facteurs explicatifs. Le modèle est testé avec des données input-output canadiennes de 1974, et il est aussi comparé à d'autres modèles.

A model of interregional freight flows for Canada. — We present in this article a model of freight flows between eight regions in Canada, and for 64 categories of goods. The model works in two stages. First, regressions are made of observed flows on some socio-economical variables, such as transportation costs. We then get prior flows, which can be modified when the explanatory variables are modified. Now, these prior flows do not necessarily respect the industrial structure of each region. To make the corrections, we resolve a mathematical program, whose objective function is based on information theory. We then look for flows as near as possible to prior flows, but which are also consistent with the industrial regional structure. This structure is included in the model using input-output accountable constraints. The model can be seen as an input-output interregional model, where the interregional coefficients are sensitive to variations in transportation costs. The formulation of the model is very flexible and allows to take into consideration other explanatory factors. The model is implemented with canadian input-output data for 1974, and compared to other models.

1. INTRODUCTION

L'étude de la demande de transport des marchandises est importante pour qui s'intéresse à la conception ou à la gestion des réseaux et systèmes de transport. Les techniques pour estimer cette demande sont cependant peu développées si on les compare aux techniques utilisées pour prévoir la demande de transport des personnes. Le peu de données disponibles à un niveau de désagrégation suffisant, tant au niveau de l'espace que des biens, et la complexité du problème expliquent en grande partie cette situation.

L'objectif que nous nous sommes donnés est de réaliser pour le Canada un modèle de prévision des flux de marchandises qui utiliserait au maximum les données disponibles. Dans un premier temps, et c'est l'objet de cet article, nous nous sommes intéressés à l'estimation des matrices origine-destination *(O-D)* de flux de marchandises, sans traiter du choix du mode de transport et de l'affectation de ces flux sur les liens du réseau de transport canadien. Il eut été possible de formuler un modèle général résolvant simultanément ces différents problèmes, mais la tâche de l'appliquer et de le rendre opérationnel avec des données empiriques, nous semblait pour l'instant trop ambitieuse.

Notre approche s'inspire des modèles input-output interrégionaux, qui permettent de représenter les liens interindustriels et les liens interrégionaux dans une économie. Ces modèles, en plus de supposer la stabilité des coefficients techniques régionaux, de façon analogue au modèle national de Léontief, supposent la stabilité des coefficients commerciaux[1]. C'est le cas, par exemple, du modèle input-output interrégional de Statistique Canada[2]. Le défaut principal de ce type d'approche, dans le cadre d'une application reliée au transport, est son insensibilité à des variations des coûts de transport entre les régions. Ils ne permettent en effet aucune substitution spatiale entre les sources d'approvisionnement des inputs.

Léontief et Strout (1963) ont proposé une version gravitaire du modèle input-output interrégional, qui tenait compte de la distance entre les régions, sans pour autant permettre de substitution sur la base de changement des coûts de transport. Wilson (1970) récrit ce modèle sous la forme d'un programme mathématique, où la fonction objectif de type entropique à maximiser est assujettie à des contraintes comptables input-output par région et à des contraintes sur les dépenses totales de transport par produit. Il présente ce modèle comme une variante des modèles d'interaction spatiale qui, dans leur présentation classique, incluent des contraintes sur l'offre et la demande des différents produits par région, sans tenir compte des interrelations entre eux.

1. Un coefficient commercial exprime la proportion d'un input nécessaire à la production d'un bien dans une région, qui provient d'une région donnée.

2. Voir Statistique Canada (1979).

Le modèle TOMM-2[3] proposé se distingue du modèle précédent, d'abord par sa façon de traiter les coûts de transport. Il ne sera en effet plus nécessaire d'estimer les dépenses totales de transport pour chaque scénario, puisqu'elles seront déterminées de façon endogène par le modèle. Il suffira donc de connaître les coûts de transport entre chaque paire *O-D*, sans pour autant connaître *ex ante* le coût total de transport. Il sera par ailleurs possible de traiter de la même façon n'importe quel autre facteur explicatif des flux interrégionaux, sans devoir estimer pour chacun la valeur totale *ex ante* devant se trouver dans le membre droit de la contrainte.

Le modèle fonctionne en deux étapes. Dans un premier temps, on régresse les flux de marchandises sur les valeurs observées des variables économiques suivantes : le coût de transport entre les régions, les exportations vers les autres régions et les importations en provenance des autres régions du Canada. On obtient de cette première étape des coefficients estimés qui permettront de prédire les flux de marchandises « a priori », étant donné des valeurs quelconques pour les variables explicatives. Ce sont des flux « a priori » puisqu'ils ne satisfont pas nécessairement la structure intersectorielle dans chaque région. Dans une deuxième étape, on corrigera donc ces flux en utilisant un programme mathématique qui est assujetti à des contraintes représentant la structure comptable input-output de chaque région. La fonction objectif de ce programme, qui repose sur le concept d'information, mesure la distance entre une distribution de probabilité a priori et la distribution a posteriori résultant du programme mathématique. En la minimisant, on s'assure donc de trouver les flux qui, tout en respectant les contraintes comptables input-ouput, sont les plus près des flux prédits par le modèle d'estimation de la première étape.

La formulation du modèle est très souple puisque plusieurs de ses éléments peuvent être modifiés, soit de façon séparée ou de façon simultanée. On peut d'abord agir au niveau du calcul des flux a priori en modifiant les valeurs des variables explicatives. Par exemple, si on modifie un ou plusieurs coûts de transport, les flux a priori s'ajustent dans un premier temps, et dans un deuxième temps, ils sont corrigés par le programme mathématique qui assure ainsi que les contraintes comptables input-output sont respectées. On peut également intervenir au niveau des contraintes comptables en modifiant, par exemple, une ou plusieurs demandes finales régionales. Dans ce cas, les flux a priori ne sont pas modifiés, mais on re-solutionne le programme mathématique avec les nouvelles contraintes. Il est finalement possible de combiner des modifications sur les variables expliquant les flux a priori et sur les contraintes, de façon à mesurer l'effet de ces modifications sur les flux interrégionaux et sur la production régionale.

Dans la deuxième section de cet article, on précisera quelle est la formulation du modèle, comment il se justifie et comment il se compare aux autres modèles qui ont déjà été proposés. Les résultats obtenus avec le modèle, des comparaisons avec d'autres modèles et les champs d'application du modèle, seront présentés

3. Pour Transportation Oriented Multiregional Model-2ième version. La première version est de Los (1980).

à la section trois. On conclura finalement sur les perspectives de développement et les problèmes soulevés par cette approche.

2. FORMULATION DU MODÈLE

Définissons d'abord la notation suivante :

$i, j = 1,\ldots,R$	représentent les régions du Canada ;
$k, h = 1,\ldots,K$	représentent les biens et services ;
B	est l'ensemble des indices représentant les biens[4] ;
x^k_{ij}	est le flux domestique[5] (en valeurs) de bien k entre les régions i et j ;
z^k_{ij}	est une valeur « a priori » pour le flux domestique de bien k entre les régions i et j ;
c^k_{ij}	est le coût de transport pour transporter un dollar de bien k entre les régions i et j ;
a^{kh}_i	est un coefficient technique domestique en valeur qui représente le montant d'un bien k d'origine canadienne utilisé comme input dans la production d'un dollar de bien h dans la région i ;
d^k_i	est la valeur de la demande finale totale de bien k dans la région i.

2.1 *L'estimation des flux « a priori »*

Dans une première étape, les flux observés de marchandises sont régressés sur les valeurs observées des variables explicatives. Chaque catégorie de biens est considérée séparément et fait l'objet d'une estimation différente. Pour chaque régression on aura ici 64 observations, soit une pour chaque paire *O-D* de la matrice de flux désagrégée en huit régions[6]. Il serait possible de proposer une spécification différente pour chaque catégorie de biens mais, comme le but est de mesurer les capacités du modèle général TOMM-2, on a préféré utiliser une spécification commune pour toutes les catégories. En plus d'une constante, les variables explicatives retenues sont le coût de transport entre les régions i et j, les importations nettes en provenance des autres régions du Canada et les exportations nettes en provenance des autres régions du Canada.

Les flux observés de marchandises proviennent des données input-output interrégionales compilées par Statistique Canada pour 1974. Ce sont de véritables flux *O-D* dans la mesure où la région d'origine est celle où est produit le bien, et la région de destination est celle où il est consommé, soit comme input

4. Donc à l'exclusion des services qui, contrairement aux biens, ne sont pas transportés sur le réseau.

5. On entend par domestique un flux dont l'origine et la destination sont canadiennes.

6. Les régions correspondent aux provinces canadiennes, le Nouveau-Brunswick, la Nouvelle-Écosse et l'Île-du-Prince-Édouard étant regroupés en une seule région.

intermédiaire ou comme consommation finale. Il n'y a donc pas de flux en transit parmi ces flux interrégionaux, qui sont donnés en valeur (en dollars et non en unités physiques). Les importations et exportations nettes vers les autres régions du Canada ont été obtenues pour chaque région et pour chaque catégorie de biens à partir de ces flux interrégionaux. L'inclusion de ces deux variables a pour but de tenir compte de certains facteurs d'inertie dans le système, en particulier la localisation des centres de production et des marchés. Finalement, les coûts de transport ont été calculés à partir d'une fonction de tarif estimée pour les chemins de fer au Canada en 1972, par Heaver et Oum (1976). Il s'agit d'une fonction log-linéaire qui explique le tarif entre deux points en fonction de la distance entre ces deux points et de caractéristiques propres à chaque catégorie de biens (densité, valeur). L'utilisation de cette fonction permet d'obtenir une première approximation des coûts de transport, qu'il sera évidemment possible d'améliorer dans une étape ultérieure de cette recherche.

Une forme fonctionnelle linéaire a été retenue pour chaque régression, et on a utilisé le modèle « tobit » de Tobin (1958) pour estimer les paramètres de ces régressions. L'utilisation de ce modèle se justifie par la présence de nombreux flux nuls pour la plupart des catégories de biens[7]. Cette concentration à une limite inférieure s'explique évidemment par l'impossibilité d'avoir des flux négatifs. Une simple droite de régression ne peut donc représenter adéquatement la relation entre les flux observés et les variables explicatives. Or, si l'on ne prend que les observations pour lesquelles les flux sont strictement positifs, on ne respecte pas les conditions du modèle des moindres carrés ordinaires puisque l'espérance mathématique de la variable aléatoire n'est plus nulle. Tobin propose alors d'utiliser toutes les observations, en se servant d'une fonction de vraisemblance où l'on fait la distinction entre les observations qui sont égales à zéro et celles qui sont supérieures à zéro.

On réalise ainsi 64 régressions différentes, soit une pour chaque catégorie de marchandises, entre lesquelles ont été répartis l'ensemble des biens transportés au Canada. On obtient de ces régressions, des coefficients qui sont significatifs à un niveau de 5 % dans tous les cas pour les variables d'importations et d'exportations nettes, et dans tous les cas sauf deux (« amiante brute » et « voitures particulières ») pour la variable coûts de transport. De plus, le coefficient R^2 est supérieur à 0,7 pour 57 des 64 régressions (Bigras, 1985).

2.2 *Une mesure de distance basée sur le concept d'information*

Les flux estimés dans une première phase ne satisfont pas nécessairement les relations comptables intersectorielles dans chacune des huit régions. Il faudra donc les corriger de façon à satisfaire ces contraintes, tout en restant le plus près

7. Il y a en moyenne 25,3 flux nuls sur 64 par matrice *O-D*.

possible de ces flux estimés. Le critère de proximité choisi est la mesure d'information développée par Kullback (1959), soit :

$$I(p,q) = \sum_{i=1}^{n} p_i \ln (p_i / q_i) \tag{1}$$

où p_i est une distribution de probabilité a posteriori et q_i est une distribution de probabilité a priori. $I\,(p,\,q)$ mesure le gain d'information de la distribution a posteriori p_i, par rapport à la distribution a priori connue q_i. Elle caractérise par le fait même la proximité entre les deux distributions de probabilité[8].

Dans le cas qui nous intéresse on peut écrire la mesure d'information en termes des flux de marchandise, soit :

$$I(x,z) = \sum_{i} \sum_{j} \sum_{k \in B} x_{ij}^{k} \ln (x_{ij}^{k} / z_{ij}^{k}) \tag{2}$$

où z_{ij}^{k} est le flux a priori estimé à partir de la régression, et x_{ij}^{k} est le résultat a posteriori. Cela est possible dans la mesure où la somme totale des flux de marchandises est donnée, soit :

$$\sum_{i} \sum_{j} \sum_{k \in B} x_{ij}^{k} = T \tag{3}$$

où T est donné de façon exogène au problème.

Si on minimise (2) en l'assujettissant à des contraintes, on s'assure alors que les flux x_{ij}^{k} obtenus s'éloignent le moins possible des flux a priori z_{ij}^{k}, tout en respectant cependant les contraintes. Parmi les flux qui satisfont à ces contraintes, le programme mathématique donnera donc les flux qui sont les plus près possibles des flux a priori estimés.

2.3 *Les contraintes intersectorielles régionales*

Chaque région est caractérisée par une structure industrielle que les données de type input-output permettent de caractériser. Au Canada, ces données sont compilées selon un cadre comptable « rectangulaire » qui distingue les biens et services des industries. Ainsi un même bien ou service peut être produit par plus d'une industrie, et une industrie peut produire plus d'un bien ou service. Ces données sont d'abord compilées au niveau national, puis désagrégées sur une base provinciale. Nous avons utilisé les données de 1974 qui étaient les dernières disponibles au moment de la réalisation de la recherche. Le niveau d'agrégation retenu est le suivant : 62 secteurs productifs ou industries, 88 catégories de produits (64 biens et 24 services), 10 catégories de la demande finale et 7 catégories de facteurs primaires.

Comme on l'a souligné précédemment, on distingue les biens, qui sont transportés sur le réseau de transport, des services qui ne le sont pas. En faisant

8. Ce que Theil (1967) a montré.

l'hypothèse que les services sont produits et consommés à l'intérieur de chaque région, il devient alors possible de relier linéairement la production de services à la production des biens, par le biais des coefficients input-output. L'avantage de cette endogénisation des services est essentiellement de réduire le nombre de variables et de contraintes du programme mathématique que l'on devra résoudre[9].

Les relations intersectorielles seront représentées en termes des biens plutôt qu'en termes des industries. Cette façon de procéder s'impose dans la mesure où nous nous intéressons aux flux de marchandises et non aux flux (parfois hétérogènes) émanant d'une industrie. La structure input-output dans chaque région impose alors les équations comptables suivantes:

$$\sum_j x_{ji}^k = \sum_{h \in B} a_i^{kh} \sum_j x_{ij}^h + d_i^k \qquad \begin{array}{l} i = 1,\ldots,R \\ k \in B \end{array} \qquad (4)$$

Ce qui arrive de bien k dans la région i, en provenance de toutes les régions du Canada (y compris de la région elle-même), est égal à ce qui est utilisé de bien k dans la région i pour produire les différents biens (demande intermédiaire), plus ce qui sert à satisfaire la demande finale de bien k dans la région i.

Wilson (1970) a souligné que cette façon de représenter la structure input-output des régions est caractéristique du modèle de Léontief et Strout (1963). L'hypothèse de base de cette approche est que la destination ultime des biens est indifférente aux producteurs, et leur origine est indifférente aux consommateurs. On peut donc concevoir un pool régional fictif pour chaque bien et dans chaque région, par lequel transitent tous les flux qui arrivent et sont utilisés dans une région.

Ces contraintes sont par ailleurs exprimées uniquement pour les biens, mais y sont inclus les quantités d'inputs nécessaires pour produire les services. En effet, la demande intermédiaire de bien k représente les besoins domestiques totaux de bien k par dollar d'output de bien h dans la région i. Cette demande intermédiaire est soit directe, soit induite par la production des services nécessaires à la production de h. Quant à la demande finale d_i^k, elle est composée de trois éléments : la production de bien k générée (directement ou par le biais des services) par les exportations des différents biens, la production de bien k induite par la demande finale des services, et la demande finale proprement dite de bien k dans la région i[10].

2.4 *Le programme mathématique*

Le programme mathématique à résoudre consiste donc à minimiser la mesure d'information (2) assujettie aux contraintes comptables input-output dans chaque région, soit les contraintes (4), et la contrainte (3) sur le total des flux dans le

9. Bigras (1985) montre que cette hypothèse n'est pas vraiment restrictive, les flux de service étant peu importants ou compensés par des flux en sens inverse.

10. On peut trouver dans Bigras (1985) le calcul détaillé de ces coefficients.

système. Il est par ailleurs possible d'ajouter d'autres contraintes à ce programme. Ainsi, pour tenir compte des capacités de production dans chaque région, des contraintes d'offre par produits peuvent être imposées, soit :

$$\sum_{j} x_{ij}^{k} + X_{i}^{k} = O_{i}^{k} \qquad \begin{matrix} i = 1,\ldots,R \\ k \in B \end{matrix} \tag{5}$$

où X_i^k sont les exportations de bien k provenant de la région i et destinées à des pays étrangers. O_i^k est alors l'offre totale de bien k provenant de la région i, que l'on considère ici exogène au modèle.

Le programme mathématique donne donc les flux qui sont les plus proches possibles des flux a priori, tout en respectant les contraintes. On peut ainsi considérer que la procédure permet de contourner la difficulté qui consisterait à faire directement les estimations sous contraintes. En effet, il ne nous est pas apparu évident comment une telle estimation pouvait se faire, d'autant plus que la structure input-output du modèle impose une résolution simultanée pour les 64 catégories de biens. Cela nous semblait poser des difficultés insurmontables d'implantation, même si cela peut s'avérer théoriquement possible.

L'examen des conditions d'optimalité permet une interprétation complémentaire du modèle qui montre bien son caractère général. Le programme mathématique consiste à minimiser une fonction strictement convexe assujettie à des contraintes d'égalité linéaires. Les conditions de Kuhn-Tucker sont alors nécessaires et suffisantes pour obtenir un optimum global. En définissant par q_i^k et v les variables duales associées respectivement aux contraintes (4) et (3), on obtient la fonction de Lagrange suivante :

$$\begin{aligned} L(x, q, v) = {} & \sum_{i}\sum_{j}\sum_{k \in B} x_{ij}^{k} \ln \left(x_{ij}^{k} / z_{ij}^{k}\right) \\ & + q_{i}^{k}\Big(\sum_{j} x_{ji}^{k} - \sum_{j}\sum_{k \in B} a_{i}^{kh}\, x_{ij}^{h} - d_{i}^{k}\Big) + v\Big(\sum_{i}\sum_{j}\sum_{k \in B} x_{ij}^{k} - T\Big) \end{aligned} \tag{6}$$

La solution du programme est alors de la forme suivante :

$$x_{ij}^{k} = z_{ij}^{k} \exp\Big(\sum_{h \in B} q_{i}^{h} a_{i}^{hk} - q_{j}^{k} - 1 - v\Big) \qquad \begin{matrix} i, j = 1,\ldots,R \\ k \in B \end{matrix} \tag{7}$$

Le modèle gravitaire input-output de Wilson, inspiré de Léontief et Strout, est alors le cas particulier où

$$z_{ij}^{k} = \exp\left(- r^{k} c_{ij}^{k}\right) \tag{8}$$

r^k étant la variable duale associée à la contrainte sur les dépenses totales de transport pour le bien k.

Le modèle TOMM-2 peut donc être vu comme une généralisation du modèle de Wilson, qui permet d'incorporer d'autres variables que le coût de transport dans l'explication des flux interrégionaux de marchandises. L'équation (7) nous permet également de voir que, contrairement aux modèles input-output interrégionaux classiques, les coefficients commerciaux ne sont pas stables. En effet, toute modification aux coûts de transport entraîne une modification du flux a priori z_{ij}^k et donc du flux *ex post* x_{ij}^k. Ajoutons que x_{ij}^k sera nul lorsque z_{ij}^k l'est, et qu'il ne sera jamais négatif, d'où la redondance de contraintes de non-négativité sur x_{ij}^k.

3. RÉSULTATS EMPIRIQUES

Les résultats du problème de minimisation sous contraintes ont été obtenus grâce à un algorithme de résolution développé par Bigras (1985). Il s'agit d'une application particulière de la méthode de balancement généralisé proposée par Lamond (1980). Il faut noter qu'il s'agit d'un programme mathématique non linéaire de grande taille puisqu'il compte 4 096 variables et au moins 512 contraintes input-output et une contrainte sur la somme totale des flux.

Pour évaluer les résultats obtenus, nous utiliserons quatre mesures d'ajustement, qui caractérisent l'écart séparant deux matrices *O-D*. On se contentera par ailleurs de présenter les résultats globaux pour l'ensemble des 64 matrices, plutôt que de les présenter pour chacune d'elles séparément[11].

La première de ces mesures est l'écart quadratique moyen :

$$EQM = [\sum_i \sum_j \sum_{k \in B} (x_{ij}^k - z_{ij}^k)^2 /4096]^{1/2} \tag{9}$$

où x_{ij}^k est le flux prédit et z_{ij}^k est le flux observé en 1974, et où il y a 4 096 paires *O-D* (soit 64 matrices de 64 éléments chacune).

La deuxième de ces mesures est l'écart absolu moyen, soit :

$$EAM = \sum_i \sum_j \sum_{k \in B} | x_{ij}^k - z_{ij}^k | /4096 \tag{10}$$

L'indice de dissimilitude, proposée par Pitfield (1978), est la troisième mesure retenue :

$$D = 50 \sum_i \sum_j \sum_{k \in B} | (z_{ij}^k/Z) - (x_{ij}^k/X) | \tag{11}$$

où Z est la somme de tous les flux z_{ij}^k et X est la somme de tous les flux x_{ij}^k. Cette mesure est facile d'interprétation puisqu'elle indique le pourcentage des flux prédits qu'il faudrait réallouer pour reproduire exactement la distribution observée des flux.

Un quatrième indicateur est retenu, qui lui n'est pas basé sur une comparaison de chacun des éléments des matrices, mais sur un critère plus agrégé. C'est le pourcentage des flux intrarégionaux ($i = j$) sur l'ensemble des flux entre les régions :

---

11. Il s'agit des résultats globaux obtenus à partir des 64 matrices individuelles, et non pas à partir d'une matrice globale des flux de toutes les catégories de marchandises.

$$XII = \sum_{i=j} \sum_{k \in B} x_{ij}^k / \sum_i \sum_j \sum_{k \in B} x_{ij}^k \tag{12}$$

3.1 *Résultats du modèle TOMM-2*

La première question qui se pose est de savoir dans quelle mesure le modèle est fidèle à la réalité. Une façon de le vérifier est de résoudre le programme mathématique en prenant comme flux a priori, la valeur estimée des flux obtenue de la régression. Il est en effet important que les résultats du modèle soient aussi proches que possible des flux observés, lorsqu'aucune modification n'est apportée, soit aux contraintes, soit aux valeurs des différentes variables explicatives. Évidemment, si on utilise directement les flux observés comme flux a priori, ce qui semble raisonnable, le problème devient trivial. Mais on veut justement se donner la possibilité de relier les flux interrégionaux à diverses variables explicatives, dont le coût de transport. En contrepartie, même si on s'éloigne des flux observés en introduisant une régression à la première étape, le résultat final des deux étapes du modèle doit être aussi proche que possible de la réalité, et donc des flux observés.

On trouve un début de réponse à cette question au tableau 1, qui donne les valeurs des quatre mesures d'ajustement définies précédemment, lorsque l'on compare les 64 matrices de flux observés aux résultats :

— de la régression ;

— du problème de minimisation sous les contraintes (3) et (4), les flux a priori étant les flux estimés obtenus de la régression ;

— du problème de minimisation sous les contraintes (3), (4) et (5), les flux a priori étant les flux estimés obtenus de la régression.

TABLEAU 1
MESURES D'AJUSTEMENT : RÉGRESSION, PROBLÈME DE MINIMISATION (AVEC ET SANS CONTRAINTES D'OFFRE)

	EQM[1]	*EAM*	*D*	*XII*[2]
Flux estimés de la régression	23 967	9 078	23,7	55,3
Problème de minimisation (sans contraintes d'offre)	26 949	7 210	17,2	61,3
Problème de minimisation (avec contraintes d'offre)	14 007	4 494	10,7	65,2

1. Le flux moyen observé est 20 868
2. Le pourcentage observé de flux intrarégionaux est 69,5%.

Ce tableau indique que les résultats obtenus à la fin des deux étapes du modèle sont relativement proches des matrices *O-D* observés. En effet, l'indice de dissimilitude montre que 17,2 % des flux devrait être réalloué pour reproduire exactement la matrice de départ, lorsqu'on n'impose pas les contraintes d'offre (5).

Lorsque ces contraintes sont ajoutées, il suffirait de déplacer 10,7 % des flux pour retrouver la matrice originale. Ces résultats marquent une certaine amélioration par rapport à ceux de la régression, puisqu'il aurait alors fallu déplacer 23,7 % des flux. Les autres mesures évoluent dans le même sens, à l'exception cependant de l'écart quadratique moyen qui s'élargit lorsqu'on ajoute les seules contraintes intersectorielles aux résultats de la régression. L'explication de cette particularité réside dans les écarts accrus pour les deux plus gros flux sur l'ensemble des 64 matrices, soit les flux intrarégionaux du Québec et de l'Ontario pour les « produits manufacturés divers ». Pour ces deux flux, les résultats de la régression sont très près des flux observés (à moins de 1,5 % près), alors que le programme mathématique avec les contraintes input-output les en éloigne à environ 10 %. Or, en termes absolus, ce sont des écarts importants, de telle sorte que 11,6 % de l'écart quadratique moyen mesuré est redevable de ces deux seuls flux (sur un total de 4 096 flux).

En termes d'ajustement des matrices, les gains provenant du seul ajout des contraintes input-output sont quand même relativement faibles si on regarde les trois autres mesures. Ceci indique que la capacité du modèle de se rapprocher des flux observés dépend plutôt des variables de la régression et, aussi, de la capacité de production des régions. L'addition des contraintes d'offre amène en effet un ajustement sensiblement meilleur, bien que dans plusieurs applications dont l'horizon temporel est le long terme, on voudra justement abandonner de telles contraintes, de façon à permettre des ajustements structurels des capacités de production régionales. On peut d'ailleurs ajouter que le modèle *TOMM*-2 prend tout son sens lorsqu'on modifie certains des facteurs expliquant les mouvements de marchandises (dans les régressions et/ou dans les contraintes). Il fallait quand même s'assurer auparavant que le modèle peut reproduire assez fidèlement les matrices originales, lorsqu'aucune modification n'est apportée aux facteurs explicatifs.

3.2 *Comparaisons avec d'autres modèles*

Il est maintenant utile de s'interroger sur les possibilités de modéliser différemment les flux interrégionaux. En particulier, il est intéressant de comparer les résultats du modèle TOMM-2, aux résultats des trois modèles suivants :

i) un modèle d'inspiration gravitaire du type proposé par Wilson, et lui-même inspiré du modèle de Léontief et Strout ;

ii) un modèle linéaire de minimisation des coûts de transport, souvent utilisé pour prédire les flux interrégionaux de marchandises[12] ;

iii) un modèle de type purement entropique qui ne tient pas compte des coûts de transport et qui distribue les flux dans les matrices de la façon la plus uniforme possible, compte tenu des contraintes qui seront imposées.

---

12. Voir par exemple, Chisholm et O'Sullivan (1973), O'Sullivan et Ralston (1974) et Baranov et Matlin (1981).

Tous ces modèles peuvent s'exprimer sous la forme d'un programme mathématique à optimiser. Afin de limiter la comparaison aux seuls effets de la formulation différente de la fonction objectif, nous les avons solutionnés avec la même méthode de résolution et en prenant les mêmes contraintes (3), (4) et (5) qui ont déjà été définies pour le modèle TOMM-2.

Les fonctions objectif à optimiser sont alors les suivantes :

i) Modèle gravitaire :

$$\text{Maximiser} \left[- \sum_i \sum_j \sum_{k \in B} x_{ij}^k \ln x_{ij}^k - \gamma \sum_i \sum_j \sum_{k \in B} c_{ij}^k x_{ij}^k \right]$$

ii) Modèle linéaire :

$$\text{Minimiser} \left[\sum_i \sum_j \sum_{k \in B} c_{ij}^k x_{ij}^k \right]$$

iii) Modèle entropique pur :

$$\text{Maximiser} \left[- \sum_i \sum_j \sum_{k \in B} x_{ij}^k \ln x_{ij}^k \right]$$

De ces fonctions objectif, seule la première a été associée, et uniquement par Wilson (1970), à des contraintes input-output. Dans tous les autres modèles pouvant s'exprimer sous la forme d'un programme mathématique, on retrouve plutôt des contraintes sur l'offre et la demande de chaque région. C'est le cas de toutes les autres versions du modèle gravitaire[13], et des modèles linéaires cités précédemment. Quant au modèle entropique pur, il s'agit en fait d'un modèle de référence où on fait l'hypothèse extrême que les coûts de transport n'ont aucune influence sur les flux de marchandises. En maximisant l'entropie assujettie à des contraintes, on obtient la solution la moins biaisée compte tenu de la seule information retenue, soit celle qui se retrouve dans les contraintes[14].

Il faut par ailleurs noter que le modèle gravitaire présenté ici est différent de la formulation originale qui supposait un paramètre différent associé à chaque catégorie de biens. En effet, comme les contraintes input-output imposent une résolution simultanée pour toutes les catégories de biens, la méthode itérative de calibration des paramètres de la fonction objectif proposée par Hyman (1969) devient inopérable avec un aussi grand nombre de paramètres[15]. On n'a d'ailleurs trouvé aucune estimation empirique du modèle proposé par Wilson en 1970. Quant aux résultats obtenus avec ces différents

13. Voir, par exemple, Chisholm et O'Sullivan (1973), Nijkamp (1975), Pitfield (1978), Vermot-Desroches (1979) et Peschel (1980).

14. Voir Jaynes (1957) et Wilson (1970) pour cette interprétation de l'entropie d'une distribution de probabilité.

15. On fixe le paramètre de façon à ce que le coût total de transport associé à la solution du programme mathématique soit égal au coût total observé. Evans (1971) ayant montré que ce coût total est une fonction monotone décroissante du paramètre, en itérant on peut trouver la solution unique. Cette procédure peut difficilement s'appliquer lorsqu'on a 64 paramètres à calibrer et que tous sont interreliés.

modèles linéaires et gravitaires, de l'avis général[16] ils consacrent leur échec relatif à modéliser les flux de marchandises. La seule prise en compte de la distance ou des coûts de transport semble en effet insuffisante pour expliquer les flux interrégionaux.

Les résultats obtenus avec les différentes formulations de la fonction objectif et sans les contraintes d'offre (5) se retrouvent au tableau 2. On remarque évidemment que le modèle TOMM-2 donne des résultats nettement supérieurs aux trois autres modèles. L'indice de dissimilitude est particulièrement intéressant puisqu'il indique que seulement 17,2 % des flux avec le modèle TOMM-2 devrait être déplacé pour reproduire exactement la matrice observée, comparativement à 69 % pour le modèle entropique qui est le moins efficace. Le pourcentage de flux intrarégionaux prédit est également très significatif, puisque le modèle TOMM-2, à 61,3 %, est à moins de 9 % du pourcentage observé de 69,5 %.

Il est par ailleurs normal que le modèle gravitaire donne de meilleurs résultats que le modèle entropique, puisqu'il ajoute à ce dernier une information supplémentaire, soit les dépenses totales de transport. Malgré tout, cela est nettement insuffisant et le faible pourcentage de flux intrarégionaux prédit par celui-ci démontre que les coûts de transport sont impuissants à tenir compte de plusieurs facteurs (institutionnels, par exemple) qui biaisent les matrice *O-D* en faveur des flux intrarégionaux. À l'inverse des modèles précédents, le modèle linéaire plutôt que de disperser les flux, tente de les concentrer sur la diagonale de chaque matrice *O-D*. Ceci est facilement compréhensible, car les coûts de transport sont évidemment les plus faibles pour ces flux à l'intérieur des régions. On a en fait défini ces coûts de transport intrarégionaux comme étant nuls dans notre modèle. Dans cette version sans contrainte d'offre, le seul empêchement à n'avoir que des flux intrarégionaux vient de ce que des régions ne produisent au départ aucune quantité de certaines catégories de biens, et que les flux ont alors dû être mis à zéro à cause de l'absence de données input-output régionales.

TABLEAU 2

MESURES D'AJUSTEMENT : MODÈLES TOMM-2, GRAVITAIRE, ENTROPIQUE, LINÉAIRE (SANS CONTRAINTE D'OFFRE)

Modèles	*EQM*	*EAM*	*D*	*XII*
TOMM-2	26 949	7 210	17,2	61,3
Gravitaire	113 584	25 073	60,2	31,0
Entropique	119 907	28 724	69,0	16,2
Linéaire	53 880	13 612	31,9	98,6

16. Voir en particulier, Pitfield (1978), Vermot-Desroches (1979) et Peschel (1980).

Ajoutons maintenant les contraintes d'offre (5) aux programmes mathématiques précédents et examinons les résultats obtenus avec les différentes formulations de la fonction objectif (*cf.* tableau 3). On remarque que l'ajustement s'améliore grandement suite à l'introduction de ces contraintes. Mesurée sur l'écart quadratique moyen, l'amélioration varie entre 37 % pour le modèle entropique et 48 % pour le modèle TOMM-2. C'est donc le meilleur modèle qui connaît la meilleure amélioration. Par contre, si on regarde l'indice de dissimilitude, les modèles entropique et gravitaire connaissent la meilleure progression, tout en restant beaucoup moins efficace que le modèle TOMM-2 qui ne nécessiterait qu'un déplacement de 10,7 % des flux prédits pour reproduire exactement les flux observés.

Soulignons finalement la difficulté de comparer les résultats de TOMM-2 aux résultats des autres modèles qu'on peut retrouver dans la littérature. C'est d'ailleurs pourquoi l'on a repris les différentes formulations ayant inspiré ces modèles, pour les comparer à TOMM-2 en utilisant les mêmes données et la même méthode de résolution. Les résultats que l'on a obtenus semblent tout de même nettement supérieurs à ceux obtenus ailleurs. Par exemple, Pitfield (1978), qui donne l'indice de dissimilitude pour chacune des 30 catégories de marchandises qu'il a retenues, indique que son meilleur résultat est de 26 % pour la catégorie « autres terres et pierres ». Or pour 62 des 64 catégories de marchandises, les résultats de TOMM-2 sont supérieurs à ce meilleur résultat de Pitfield.

TABLEAU 3

MESURES D'AJUSTEMENT : MODÈLES TOMM-2, GRAVITAIRE, ENTROPIQUE, LINÉAIRE (AVEC CONTRAINTE D'OFFRE)

Modèles	*EQM*	*EAM*	*D*	*XII*
TOMM-2	14 007	4 494	10,7	65,2
Gravitaire	62 464	9 674	23,2	51,9
Entropique	76 017	16 442	39,5	32,0
Linéaire	31 853	7 597	19,2	84,3

3.3 *Champs d'application du modèle*

Le modèle TOMM-2 permet une grande variété de simulations. On peut d'abord agir au niveau des variables socio-économiques expliquant les flux a priori. Par exemple, si on diminue les coûts de transport pour tous les flux de marchandises (ou pour certains d'entre eux) qui partent d'une région, on pourra voir les effets sur les flux interrégionaux et sur la production régionale. L'impact se fait d'abord sentir au niveau des flux a priori qu'on recalcule à partir des coefficients estimés dans les régressions. Ces nouveaux flux a priori sont ensuite insérés dans le programme mathématique, dont la solution tendra vers ces nouveaux flux a priori tout en respectant cependant les contraintes qu'on leur impose.

On peut aussi agir au niveau des contraintes en modifiant certaines demandes finales régionales, tant pour les biens que pour les services. Ceci permettra de calculer des multiplicateurs intersectoriels analogues aux multiplicateurs input-output. En fait, ce sont les mêmes multiplicateurs mais ils ne sont pas obtenus en inversant une matrice mais en solutionnant un programme mathématique. Les modifications peuvent aussi porter sur les coefficients structurels intersectoriels régionaux, si on veut simuler les effets de politiques dont le but serait d'agir sur la structure industrielle de certaines provinces.

Il est finalement possible d'utiliser les matrices origine-destination données par TOMM-2 comme input dans un modèle de transport simulant le choix modal et l'affectation sur le réseau de transport multimodal canadien, à un niveau de désagrégation plus poussée. Cette application permet de relier la demande de transport des marchandises par modes aux variables économiques structurelles que sont les coefficients input-output régionaux et aux variables économiques conjoncturelles que sont les demandes finales régionales.

4. CONCLUSION

Les résultats obtenus avec le modèle TOMM-2 sont supérieurs à ceux obtenus avec les autres modèles des flux interrégionaux de marchandises qui considèrent également les coûts de transport entre les régions. La supériorité de la fonction objectif retenue est évidente par rapport aux formulations gravitaire, entropique et linéaire, dont se sont inspirés les autres modèles. Cela s'explique par le fait que le modèle inclut d'autres facteurs explicatifs que les coûts de transport, au niveau des régressions et au niveau des contraintes. Le modèle est en fait plus général que les modèles précédents.

Par rapport aux modèles input-output classiques, la comparaison ne peut évidemment se faire sur les mêmes bases. Par définition, ces derniers reproduisent exactement les données de l'année de base à partir de laquelle ils ont été calculés. Leur grande faiblesse, que TOMM-2 a voulu justement résoudre, c'est leur incapacité à permettre des substitutions interrégionales, lorsque les coûts de transport sont modifiés par exemple. Cet inconvénient est évidemment majeur dans le cadre d'applications reliées au transport.

Au niveau méthodologique, le modèle fonctionne en deux étapes séquentielles. Cette procédure nous est évidemment imposée par l'impossibilité pratique de résoudre de façon simultanée les 64 régressions sous les contraintes (3), (4) et (5). Malgré cela il nous semble justifié de percevoir une certaine rationalité économique dans cette procédure. Les flux a priori obtenus des régressions sont en quelque sorte des flux *ex ante* qui correspondent aux décisions qui seraient prises par les agents économiques en l'absence des contraintes structurelles régionales. Or, il est possible de prétendre que ces décisions se prennent effectivement sans considérer tous les ajustements que peuvent impliquer les contraintes structurelles. Les agents économiques qui augmentent leur demande pour un bien et leur demande de transport pour ce bien, ne voient pas nécessairement les effets que cela peut avoir sur la demande de transport de tous les biens intermédiaires, par exemple. C'est pourquoi la structure industrielle propre à chaque région vient *ex post* imposée les ajustements nécessaires sur ces décisions a priori des agents économiques.

La formulation du modèle est par ailleurs très souple. Ainsi des modifications peuvent être apportées aux valeurs des différentes variables explicatives des régressions ou aux différents éléments (demande finale ou coefficients structurels) composant les contraintes input-output. Ceci permet évidemment une grande variété d'applications combinant, par exemple, des modifications sur les coûts de transport et la demande finale. On peut également ajouter des informations pertinentes sous forme de contraintes, ou définir pour chaque catégorie de biens une spécification particulière à l'étape initiale de la régression des flux observés sur des variables socio-économiques appropriées. Par ailleurs, le modèle a le grand avantage sur les modèles de type gravitaire de déterminer de façon endogène les dépenses totales de transport. Il s'agit là d'un problème qui n'avait jamais été résolu avec ce type de modèle.

Le modèle n'en a pas moins certaines faiblesses et limitations. Il fonctionne à un niveau d'aggrégation assez élevé et ne traite ni du choix modal ni de l'affectation. Par ailleurs, les coûts de transport ont été approximés à partir d'une fonction de tarif estimée pour les chemins de fer, ce qui pourrait être corrigé à partir d'informations plus désagrégées, spatialement et par mode de transport. À l'intérieur du cadre général présenté dans cet article, il demeure donc possible d'apporter un certain nombre d'améliorations qui pourront enrichir le modèle TOMM-2.

BIBLIOGRAPHIE

BARANOV, E.V. et I.S. MATLIN (1981), « Systems of models coordinating decisions for sectoral and regional development », communication présentée à un colloque de l'International Institute for Applied Systems Analysis, Laxenburg (Autriche).

BIGRAS, Y. (1985), « La modélisation des flux de marchandises au Canada », thèse de Ph.D., publication 483, Centre de recherche sur les transports, Université de Montréal.

CHISHOLM, M. et P. O'SULLIVAN (1973), *Freight Flows and Spatial Aspects of the British Economy*, Cambridge University Press, London.

EVANS, A.W. (1971), « The Calibration of Trip Distribution Models with Exponential or Similar Cost Functions », *Transportation Research*, vol. 5, pp. 15-38.

HEAVER, T.D. et T.H. OUM (1976), « A Statistical Analysis of the Canadian Railway Rate Structure », *Transportation Research Forum Proceedings, 17th Annual Meeting*, vol. 19, pp. 570-577.

HYMAN, G.M. (1969), « The Calibration of Trip Distribution Models », *Environment and Planning*, vol. 1, pp. 105-112.

JAYNES, E.T. (1957), « Information Theory and Statistical Mechanics », *Physicals Review*, vol. 106, pp. 620-630.

KULLBACK, S. (1959), *Information Theory and Statistics*, Wiley, New York.

LAMOND, B. (1980), « Balancement de matrices et optimisation d'entropie », Publication n° 211, Centre de recherche sur les transports, Université de Montréal.

LEONTIEF, W. et A. STROUT (1963), « Multiregional Input-Output Analysis », in T. Barna, éd., *Structural Interdependence and Economic Development*, MacMillan, London, pp. 119-150.

LOS, M. (1980), « A Transportation-Oriented Multiregional Economic Model for Canada », publication n° 178, Centre de recherche sur les transports, Université de Montréal.

NIJKAMP, P. (1975), « Reflections on Gravity and Entropy Model », *Regional Science and Urban Economics*, vol. 5, pp. 203-225.

O'SULLIVAN, P. et B. RALSTON (1974), « Forecasting Intercity Commodity Transport in the U.S.A. », *Regional Studies*, vol. 8, pp. 191-195.

PESCHEL, K. (1980), « On the Impact of Geographic Distance on the Interregional Patterns of Production and Trade », Discussion paper n° 15, Institut for Regionalforschung, Université de Kiel.

PITFIELD, D.E. (1978), « Freight Distribution Model Predictions Compared : a Test of Hypothesis », *Environment and Planning A*, vol. 10, pp. 813-866.

STATISTIQUE CANADA (1979), *Canadian Interregional Input-Output Tables — Sources and Methodology*, Statistique Canada, Division de l'analyse structurelle, Ottawa.

THEIL, H. (1967), *Economics and Information Theory*, North Holland, Amsterdam.

TOBIN, J. (1958), « Estimation of Relationships for Limited Dependent Variables », *Econometrica*, vol. 26, pp. 24-36.

VERMOT-DESROCHES, B. (1979), « Testing Econometric Spatial Interaction Models Using French Regional Data », Communication présentée à la 26e rencontre nord-américaine de la Regional Science Association, Los Angeles.

WILSON, A.G. (1970), *Entropy in Urban and Regional Modeling*, Pion, London.