

## Les échanges bilatéraux entre les nations : Une approche linéaire mixte des modèles gravitationnels\*

*Kamel Ghaddab*<sup>1</sup>

*Ahmed Silem*<sup>2</sup>

### Introduction

Dans le cadre de la détermination empirique de la composition et du potentiel des flux des échanges commerciaux, l'approche empirique gravitaire a suscité de nombreux travaux dans les dernières décennies. Le questionnement central posé dans le cadre de la littérature économique était de développer un soubassement théorique à cette lignée de recherches empiriques. Il était question de dériver une forme structurelle de la dite « équation gravitaire » du commerce bilatéral entre les nations. La forme ainsi que la méthode d'estimation ont constitué un champ d'investigation empirique controversé, et ce en matière de la technique économétrique appropriée et de l'interprétation des résultats obtenus. La controverse a aussi porté sur l'identification de la variable libéralisation commerciale.

Trois essais sont représentatifs de ces travaux. Frankel et Romer [1999] se sont penchés sur la détermination empirique d'un degré d'intégration tel qu'il est estimé de l'équation gravitaire. L'instrument établi par les auteurs a fait l'objet d'une vaste utilisation dans tous les travaux portant sur la détermination empiriques du développement économique, ou dans le cadre des recherches sur les causalités complexe liant le développement à ses déterminants dits « profonds » à long terme. Wei [2000], d'une manière similaire aux travaux de Frankel et Romer [1999], introduit la notion d'intégration « naturelle » comme approximation pour la valeur prédite à partir de l'équation gravitaire pour le volume des échanges bilatéraux entre les économies. Rose [2002] propose un instrument basé sur les erreurs estimées de l'équation gravitaire pour fournir une approximation du degré du protectionnisme des économies.

---

\* la communication proposée correspond à une partie de la recherche entreprise dans le cadre de la thèse en doctorat ès sciences économiques par Kamel Ghaddab, sous la direction de Ahmed Silem.

<sup>1</sup>.Institut supérieur de gestion de Sousse, équipe de recherche Euristik, Centre Magellan Université Jean Moulin Lyon 3, kamelghad@yahoo.fr  
<sup>2</sup>Equipe de recherche Euristik, Centre Magellan Université Jean Moulin Lyon 3, silemahmed2@gmail.com

L'hypothèse qui sous tend ce travail étant que l'équation gravitaire fait intervenir les variables relatives à la taille, au niveau de développement et à la composante géographique des échanges. Le résidu peut être imputé aux différentes distorsions à l'échange des économies, qui sont engendrées par les politiques économiques, particulièrement celle de nature commerciale.

Dans le cadre du présent travail, l'hypothèse de Rose [2002] sera retenue mais amendée par une approche alternative pour l'estimation de l'équation gravitaire. En effet, et étant donné l'importance des erreurs dans le cadre de ces recherches, la technique ainsi que l'approche d'estimation pourraient jouer un rôle déterminant dans la précision de l'instrument. L'application des moindres carrés ordinaires (MCO) dans le cas d'une équation gravitaire avec hétérogénéité des coefficients entre les paires pourrait engendrer une absence de précision des écarts types estimés pour ces coefficients, et donc une perte conséquente de la consistance des estimateurs, donc de l'erreur estimée. L'approche alternative proposée est celle des modèles linéaires mixtes, qui sont plus appropriés dans le cas d'une hétérogénéité des paires en termes de leurs volumes moyens des échanges, et de leurs coefficients reflétant l'effet de la taille et du niveau de développement économique.

Le travail sera développé en deux parties. Dans la première, il sera question d'étudier les limites objectives des techniques traditionnelles d'estimation de l'équation gravitaire des échanges. La deuxième partie sera une application sur données de panel.

## Section 1 Approches Empiriques des Modèles Gravitationnels

Le développement de la nouvelle économie géographique (NEG) ces dernières années a suscité un intérêt renouvelé à l'égard des modèles gravitationnels pour la détermination du volume des échanges bilatéraux entre les nations. L'intérêt pour la NEG est justifié généralement par le fait que la modélisation des rendements croissants et des formes de la concurrence imparfaite a permis la réintégration de l'espace dans la théorie du commerce international<sup>3</sup>. L'équation gravitaire permet d'estimer dans ce cas les divers canaux de transmission des effets de la géographie et des

---

<sup>3</sup> Les travaux pionniers sont ceux de Ohlin Bertil (1933) *Interregional and International Trade*, Balassa Bela (1967), Kravis I. B., R. E. Lipsey. (1971) *Price Competitiveness in World Trade*, UMI, NBER, 1971, Dixit A., J. Stiglitz (1977) "Monopolistic Competition and Optimum Product Diversity", *AER* 67, Krugman Paul (1979-1990 - 1997), *Development, Geography, and Economic Theory*. MIT Press

interdépendances économiques dans l'espace sur le volume du commerce bilatéral entre les nations. Néanmoins, plusieurs problèmes quantitatifs et économétriques rendent la tâche délicate.

C'est ainsi que le problème de quantification de certaines variables économiques prédites théoriquement en tant qu'arguments du volume des échanges oblige à opérer un ensemble d'approximations statistiques. Ce problème concerne principalement les variables relatives aux diverses barrières à l'échange telles que les tarifs bilatéraux et les coûts de transport, d'une part, et surtout celles relatives aux indices de prix prédits théoriquement, d'autre part.

Dans le cadre de la littérature empirique, deux approches ont été développées comme cadre d'estimation des équations gravitaires. Dans une première vague de recherches empiriques, les spécifications estimées pour l'équation gravitaire ont été quasi entièrement basées sur des données en coupe transversale et en appliquant la technique des moindres carrés ordinaires. Ce n'est que dans les années quatre vingt dix que l'approche d'estimation sur données de panel a été introduite, et pour devenir l'approche dominante en cette matière<sup>4</sup>.

### 1. Structure générale des estimations sur données de panel :

La spécification générale d'un modèle en données de panel est donnée par :

$$y_{it} = \sum_{k=1}^k x_{kit} \beta_{kit} + \varepsilon_{it}, \quad i = 1 \dots N, t = 1 \dots T \quad (1)$$

Où  $x_{it}$  représente un vecteur  $1 \times k$  relatif aux variables indépendantes disposant d'une variabilité inter temporelle et interindividuelle,  $\beta_k$ , étant le vecteur des coefficients respectifs à ces variables. En dissociant l'ensemble des variables explicatives dépendantes du temps et celui des variables ne variant que sur le plan interindividuel, et en notant par  $z_i$  ce dernier ensemble, on peut dériver l'équation suivante :

$$y_{it} = x_{it} \beta_k + z_i \delta + u_i + \varepsilon_{it} \quad (2)$$

$u_i$ , représente l'effet spécifique aux individus. Il est supposé que ce terme est non corrélé avec le terme d'erreur  $\varepsilon_{it}$ , mais concernant sa corrélation avec les variables indépendantes  $x_{it}$  et  $z_i$ , on distingue deux types de modèles économétriques distincts : les modèles à effets aléatoire dans le cas d'une corrélation entre l'effet individuel et le vecteur des variables indépendantes, et les modèles à effet fixe dans le cas contraire.

---

<sup>4</sup> Voir Harris et Matyas [1998] pour une discussion technique approfondie des spécifications économétriques de l'équation gravitaire sur données de panel et en coupe transversale.

Dans un premier temps, on isole la spécification des effets fixes qui peuvent être étudiés selon les deux dimensions temps et espace. Concernant la pente, elle est toujours supposée commune à toutes les unités transversales et elle est constante dans le temps. Cette variante du modèle est qualifiée souvent du modèle des moindres carrés avec variables indicatrices (LSDV).

Formellement, on peut adopter la transformation courante qui consiste à évaluer les variables par rapport à la moyenne par individu.

$$\bar{y}_{it} = \bar{x}_{it} \beta + \bar{\varepsilon}_{it} \quad (3)$$

L'estimateur, qualifié de  $\beta_{Fixe}$ , dérivé de cette spécification mesure l'effet de la déviation d'une variable explicative par rapport à la moyenne sur la déviation de la variable dépendante par rapport à la moyenne. Dans notre cadre empirique, la limite de cette spécification est que les effets des variables explicatives indépendantes du temps ne vont être estimés qu'à travers la constante spécifique aux individus.

Une deuxième manière d'estimer les effets fixes consiste à utiliser les moyennes de l'échantillon global comme vecteur d'observations, et cela pour la variable endogène et les variables explicatives. La spécification de base de ce modèle est donnée par l'équation suivante :

$$\bar{y}_i = \bar{x}_i \beta + \bar{z}_i \delta + u_i + \bar{\varepsilon}_i \quad (4)$$

L'estimateur dans ce cas est qualifié de la  $\beta_{Between}$

Dans le cas de l'hypothèse que cette spécificité individuelle est aléatoire et supposée indépendante des variables explicatives, on retrouve la spécification des effets aléatoire.

$$y_{it} = x_{it} \beta + z_i \delta + (u_i + \varepsilon_i) \quad (5)$$

Cette spécification permet d'estimer l'hétérogénéité spécifique aux individus. L'estimateur dérivé de cette spécification est qualifié de la  $\beta_{Random}$ .<sup>5</sup>

## 2. Fondements économétrique et structure générale de la spécification hiérarchique :

Spécification de l'équation gravitaire au niveau 1 :

$$LTrade_{tij} = \beta_{0ij} + \beta_{1ij} LGdp_{tij} + \beta_{2ij} LGdppc_{tij} + \varepsilon_{tij} \quad (6)$$

---

<sup>5</sup> Voir Rabe-Hesketh et Everitt Bian. [2007] pour une discussion détaillée sur les différentes formes des modèles sur données de panel ainsi que sur leur application sur Stata.

Avec  $\beta_{qij}$  les coefficients associés aux variables explicatives du niveau 1.  $\sigma^2$ , représente l'effet aléatoire du premier niveau, c'est-à-dire du terme aléatoire  $\varepsilon_{tij}$ , supposé suivre une loi normale  $\varepsilon_{tij} \rightarrow N(0, \sigma^2)$ .

Spécification de l'équation gravitaire au niveau 2 :

$$\beta_{0ij} = \gamma_{00} + \gamma_{01}Ldist_{ij} + \gamma_{02}Border_{ij} + \gamma_{03}Colony_{ij} + \gamma_{04}Comcol_{ij} + \gamma_{05}Landl_{ij} + \mu_{00} \quad (7)$$

$$\beta_{1ij} = \gamma_{10} + \mu_{10} \quad (8)$$

$$\beta_{2ij} = \gamma_{20} + \mu_{20} \quad (9)$$

$$LTrade_{tij} =$$

$$\gamma_{00} + \gamma_{01}Ldist_{ij} + \gamma_{02}Border_{ij} + \gamma_{03}Colony_{ij} + \gamma_{04}Comcol_{ij} + \gamma_{05}Landl_{ij} + \gamma_{10}LGdp_{tij} + \gamma_{20}LGdppc_{tij} + \mu_{10} * LGdp_{tij} + \mu_{20} * LGdppc_{tij} + \varepsilon_{tij} + \mu_{00} + \mu_{10} + \mu_{20}$$

(10)

Le terme composé ( $\mu_{10} * LGdp_{tij} + \mu_{20} * LGdppc_{tij} + \varepsilon_{tij} + \mu_{00} + \mu_{10} + \mu_{20}$ ) représente la composante stochastique du modèle.

Au deuxième niveau du modèle, on suppose que le vecteur  $\mu_{qij} = \begin{pmatrix} \mu_{q1} \\ \mu_{qN} \end{pmatrix}$  suit une loi normale

multivariée de moyenne nulle et de variance  $\tau_{qij}$ .

Pour chaque paire d'effets aléatoires  $q$  et  $q'$ , la covariance est supposée :  $Cov(\mu_{qij}, \mu_{q'ij}) = \tau_{qq'}$

La forme particulière classique de l'équation gravitaire (1) matérialise le modèle classique des MCO. Dans ce modèle, un seul facteur aléatoire est retenu, dont la variance est supposée refléter la dispersion globale de l'échantillon. Une des hypothèses fondamentales de ce modèle est l'indépendance du terme  $\varepsilon_{tij}$ ,  $var(\varepsilon) = I \sigma^2$ . Cette hypothèse est de nature à supposer implicitement l'indépendance des variables explicatives  $Y_i$ .

Dans le cadre de cette nouvelle lignée de modèles, la forme classique est qualifiée d'un modèle fixe dans le sens où tous les facteurs sont supposés fixes, à l'exception de l'unique facteur aléatoire matérialisé par le terme d'erreur classique.

Ces deux familles de modèles - à effets fixes et à effets aléatoires - sont complétées par la lignée de modèles qualifiés de mixtes, car ils tiennent compte de deux types d'effets. Etant donné la nature complexe de la définition de ces effets, il serait intéressant de retenir les définitions communément partagées dans le cadre de la littérature en cette matière. D'abord, ces effets sont définis lorsque l'échantillon objet d'étude est de nature stratifiée ou groupée. Dans ce sens, les effets fixes sont définis comme étant les paramètres associés à la population totale, et dont les variables en question sont observables d'une manière répétitive pour tous les individus. Par opposition, on définit les facteurs aléatoires par les paramètres associés aux unités prises comme sous échantillon aléatoire de la population totale objet d'analyse. Dans ce modèle, on peut expliquer une variable dépendante par des variables dites «covariantes» qui sont classées par des facteurs matérialisant la stratification de la population totale. Dans ce cas, cette approche permet une analyse de la structure de la variance globale et de la covariance entre les termes aléatoires du modèle.

La structure générale du modèle linéaire mixte peut être donnée par :

$Y = X\beta + Z\mu + \varepsilon$ , avec  $X_{N \times p}$  est une matrice de constantes connues,  $Z_{N \times q}$  est une matrice de constantes connues,  $\beta$  représente le vecteur des paramètres associés aux effets fixes,  $\mu$  représente le vecteur des paramètres associés aux effets aléatoires et  $\varepsilon$  représente le terme aléatoire classique.

On suppose que les termes aléatoires du modèle suivent des lois normales :

$\begin{pmatrix} \mu \\ \varepsilon \end{pmatrix} \rightarrow N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right)$  Avec G et R qui sont des matrices de formes connues mais qui dépendent d'un ensemble de paramètres inconnus  $\theta$ .

Dans ce cas, la variance de la variable dépendante Y est donnée par :  $var(Y) = ZGZ' + R$  qui est différente de  $I\sigma^2$ , qui est de nature à remettre en cause l'efficacité des estimateurs de la méthode des MCO. Pour relever cette indétermination, il est donc plus pertinent d'employer l'estimateur des moindres carrées généralisées qui consiste en effet à minimiser le terme  $(Y - X\beta)' var_Y^{-1} (Y - X\beta)$ . Notons que la variance de la variable dépendante n'est plus connue, mais par sa dépendance des paramètres  $\theta$ , il faudra estimer ces derniers pour déterminer la valeur estimée des coefficients  $\beta$ .

La méthode d'estimation généralement utilisée dans cette lignée de modèles linéaires mixtes est la méthode de maximum de vraisemblance. Sous l'hypothèse de normalité des erreurs, cette méthode consiste à dériver un estimateur  $\hat{\theta}$  de  $\theta$ , et ce en minimisant le terme de déviation :  $-2 \text{Log } L(\theta) = \text{Log } |V_Y| + N \text{Log } (\hat{\epsilon}' V_Y^{-1} \hat{\epsilon})$ , avec  $\hat{\epsilon} = Y - X (X' V_Y^{-1} X)^{-1} X' V_Y^{-1} Y$ . L'estimateur dérivé de cette minimisation est qualifiée d'estimateur de maximum de vraisemblance noté  $\beta_{ML}$ .

## Section II : Estimation d'un modèle linéaire mixte de l'équation gravitaire

### -A. Données, définition des variables et formes particulières des spécifications :

#### A-1. Présentation des variables :

L'échantillon global est composé de 116 pays dont la liste est décrite en annexe. Sur la base de cette population totale, on définit une sous population aléatoire de 2170 paires d'échangistes.

Les données objet d'analyse sont observées sur la période 1980-2004. Les données sont relatives aux variables suivantes :

Variables du premier niveau de l'estimation :

- $Ltrade_{tij}$  : cette variable mesure la valeur réelle (prix constants 2000) des échanges bilatéraux entre chaque paire d'échangistes.
- $Lgdp_{tij}$  : cette variable mesure le produit logarithmique des produits intérieurs bruts réels des deux partenaires composant une paire.
- $Lgdppc_{tij}$  : cette variable mesure le produit logarithmique des produits intérieurs bruts réels per capita des deux partenaires composant une paire.

Variables du second niveau de l'estimation :

- $Ldist_{ij}$  : cette variable représente le niveau logarithmique de la distance géographique bilatérale séparant les deux pays d'une paire.
- $Landl_{ij}$  : cette variable représente le nombre de pays enclavé géographiquement par paire. Elle peut prendre la valeur : 0, 1 ou 2.

- $Comlang_{ij}$  : cette variable représente une variable binaire reflétant le fait que les pays de chaque paire partagent ou non la même langue.
- $Colony_{ij}$  : cette variable est une variable binaire, elle reflète le fait que les pays de chaque paire partagent ou non les mêmes historiques coloniaux.
- $Border_{ij}$  : cette variable binaire reflète la présence ou non d'une frontière géographique commune entre les pays d'une même paire.

#### *A-2. Sources des données et explication des approximations statistiques :*

Concernant la variable relative au volume moyen des échanges des paires, la source brute des données correspond aux travaux statistiques d'Alessandro Nicita et Marcello Olarreaga [2006]. La base initiale est constituée de plus de 5 millions observations relatives au nombre maximum de pays et de paires pour lesquels l'ensemble ou une partie des données sur cette variable est disponible. Les séries brutes de données sont très discontinues dans le temps<sup>6</sup>. Concernant la composante taille de l'équation gravitaire, le choix d'intégration des variables a constitué une diversité, voire une divergence dans la littérature empirique en cette matière. Cette divergence s'explique en grande partie par le cadre empirique dans lequel cette équation gravitaire est spécifiée. Généralement, le produit intérieur brut, la population totale et la taille physique des nations sont souvent retenues comme approximation de l'effet de taille communément prédit dans les formes structurelles de l'équation gravitaire. En effet, la corrélation théorique prédite entre la variable population totale et celle relative au volume des échanges bilatéraux est négative, toutes choses étant égales par ailleurs, c'est-à-dire après avoir tenu en compte l'effet du produit intérieur brut comme approximation du niveau de développement économique des nations en question.

Pour ce qui est du produit intérieur brut (ou alternativement le produit national brut), son utilisation ainsi que l'interprétation de son signe (et même son amplitude) ont fait l'objet d'une large controverse. Cette controverse s'explique par le fait que l'équation gravitaire a été généralement utilisée comme justification à des thèses théoriques généralement rivales.

---

<sup>6</sup> Le raffinement des données, le choix des pays, des paires et de la période de référence ainsi que les calculs d'agrégation des observations sur les flux du commerce bilatéral sont le fruit des travaux de Kamel Ghaddab

Les approches théoriques qui ont constitué les soubassements aux cadres empiriques divers sont principalement les théories factorielles de l'échange international, les nouvelles théories du commerce international dites de concurrence imparfaite et les théories de croissance endogène.

Enfin, et pour ce qui de la variable taille géographique des nations, la justification de son intégration dans l'équation gravitaire se base généralement sur le fait que cette variable est une approximation indirecte des dotations en termes de ressources naturelles.

Dans notre proposition, et parmi les diverses spécifications existantes, l'approche de Frankel [1997] est retenue. Elle s'est basée sur les variables PIB et PIB per capita en tant que variables reflétant l'effet taille de l'équation gravitaire. Il est possible de justifier l'indépendance de ces deux variables comme variables explicatives dans une même spécification. Bien qu'il paraisse relativement facile de trouver une justification du signe du produit intérieur brut, l'interprétation d'un effet indépendant du produit intérieur brut per capita est relativement compliquée. Cette interprétation va de la recherche et le désir de la variété pour les consommateurs d'une nation jusqu'à l'explication présentée par les nouvelles théories de la croissance et qui se base sur une explication schumpetérienne de la dynamique d'accumulation basée sur l'innovation.

Les données relatives aux variables, produit intérieur brut en niveau agrégé et per capita, sont collectées des indicateurs de développement de la banque mondiale [2006].

Pour les variables de deuxième niveau de l'estimation, et étant donné la logique de cette approche linéaire mixte de l'équation gravitaire, ces variables ne touchent que les paires et sont indépendantes du temps. Elles sont ainsi supposées expliquer la dispersion des niveaux moyens d'échange bilatéral ainsi que les pentes relatives à la corrélation entre l'échange bilatéral et les variables de taille apparaissant uniquement au premier niveau.

Comme variables spécifiques aux paires, on retient la distance bilatérale comme approximation indirecte de l'effet du coût de transport sur le volume des échanges bilatéraux pour une paire donnée. La présence ou non d'une frontière commune entre les économies composant une paire est intégrée comme deuxième variable approximant la proximité.

La justification d'un effet indépendant de cette variable est que le fait de partager des frontières communes pourrait jouer un rôle complémentaire au rôle des coûts de transport comme déterminant des flux des échanges.

Les données sur la distance bilatérale séparant les deux villes les plus importantes dans chaque nation sont collectées de la base de données d'Andrew Rose.

D'autres variables sont généralement retenues comme déterminant pour la fluidité des échanges entre les économies. Dans ce cas, c'est la proximité institutionnelle et culturelle qui est recherchée. Cette forme de proximité est jugée fonction des similarités coloniales et par les liens de langue que les nations partagent. Ces variables reflétant les aspects géographiques et institutionnels sont collectées de la base de données d'Andrew Rose.

## **-B. Estimation et Interprétations**

### **B-1. Axes généraux de la méthodologie empirique :**

Notre méthodologie empirique dans le cadre de ce travail sera développée au tour de trois axes principaux :

- a) Dans une première étape, nous procéderons à une estimation sur des données de panel des spécifications proposées pour l'équation gravitaire. La période couverte par notre analyse s'étale entre 1980 et 2004..La justification de ce type d'estimation sur données de panel comme étape préalable aux estimations de la forme linéaire hiérarchique de l'équation gravitaire est justifiée pour nos propos par le fait que cette dernière forme pourrait être présentée comme un dépassement des formes classiques sur données de panel. Ainsi, si les tests infirment la stabilité recherchée pour les coefficients, la solution plausible est de passer à une spécification hiérarchique de l'équation gravitaire.
- b) Dans une deuxième étape, les estimations seront effectuées pour toutes les spécifications de la forme linéaire mixte de l'équation gravitaire. Trois spécifications principales vont constituer notre travail, et qui sont respectivement l'estimation de la forme dite «empty» de l'équation ou ce qu'on appelle la forme d'analyse de la variance de la variable dépendante de l'équation.

La deuxième spécification aura comme objectif d'estimer le modèle en supposant que la constante du modèle est aléatoire. La constante reflète en effet le volume moyen des échanges des paires dans un modèle centré par rapport à la moyenne. En fin, une spécification du modèle sous l'hypothèse que tous les coefficients sont aléatoires sera estimée.

- c) La dernière étape consistera à effectuer tous les tests statistiques pour étudier la robustesse des résultats dégagés. Dans le même cadre, et vue l'importance cruciale de l'erreur estimée dans la forme linéaire mixte de l'équation gravitaire, un traitement de la qualité des erreurs sera effectué.

L'objectif sera de dériver un indicateur de politique commerciale comme approximation de l'erreur estimée du modèle.

## B-2. Formes de panel des échanges bilatéraux et corrélation résiduelle :

Dans ce cadre, on va étudier la forme de panel retenue par l'équation gravitaire proposée et d'estimer le degré de la corrélation résiduelle, ou ce qu'on qualifie du coefficient de corrélation intra class du modèle. L'objectif étant d'estimer une décomposition de la variance globale pour connaître la variance potentiellement expliquée par la dispersion de la variable dépendante par des aspects spécifiques aux paires d'échange et qui ne sont pas explicitée dans la dimension longitudinale des données pour l'équation gravitaire.

Le tableau 14 représente l'ensemble des statistiques descriptives sur les variables du modèle en question.

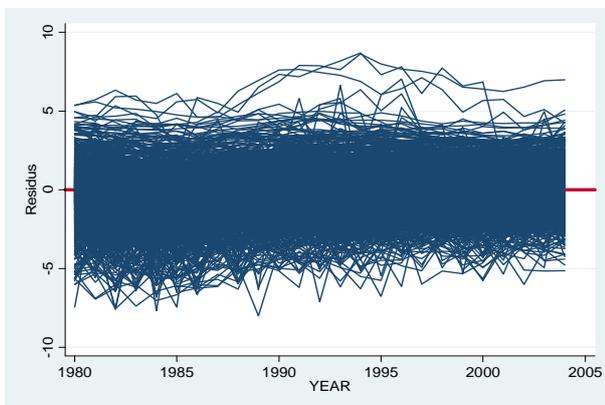
Dans un premier stade, notre objectif est de tester est ce que les résidus sont corrélées

$$y_{it} = \beta_0 + \sum_{k=1}^k x_{kit} \beta_{kit} + \varepsilon_{it}, \quad i = 1 \dots N, t = 1 \dots T$$

$$\hat{\varepsilon}_{tij} = y_{tij} - (\hat{\beta}_0 + \sum \hat{\beta}_k \hat{x}_{ktij})$$

Comme le montre la figure 1, les erreurs sur la période globale et pour une même paire évoluent autour de la moyenne nulle, donc elles sont corrélées. Dans ce cas, et même si la consistance des estimateurs des MCO ou du maximum de vraisemblance demeure garantie, il n'en est plus ainsi pour les écarts types estimés des coefficients.

Figure 1 : Corrélation résiduelle pour un échantillon aléatoire de paires 1980-2004



A ce niveau, il sera intéressant d'analyser d'une manière explicite les composantes de l'erreur ou de la variance globale. En effet l'erreur globale  $\epsilon_{tij}$  peut être décomposée en deux parties : une première partie de l'erreur spécifique aux individus  $ij$ , matérialisant la dépendance intra individuelle, dans le sens où elle est permanente et indépendante du temps  $\mu_{ij}$ , et une seconde partie transitoire et qui reflète la variabilité intra et inter individuelle  $\vartheta_{tij}$ .

Dans le cadre de cette approche linéaire mixte, l'erreur globale  $\epsilon_{tij}$  est donc décomposée en une erreur de niveau 2 qui reflète la dépendance permanente intra individuelle et donc elle peut être qualifiée de l'hétérogénéité non observable, et une seconde erreur de niveau 1 qui reflète la variabilité inter et intra individuelle de la population totale des paires.

Etant donné cette décomposition, on peut expliciter la forme primaire du modèle linéaire mixte qualifié du modèle à constante aléatoire :

$$y_{it} = (\beta_0 + \mu_{ij}) + \sum_{k=1}^k x_{kit} \beta_{kit} + \vartheta_{tij}, \quad i = 1 \dots \dots N, t = 1 \dots \dots T$$

Cette formulation permet ainsi d'explicitier les deux niveaux de la régression. En effet, le modèle permet d'intégrer une constante aléatoire  $\mu_{ij}$  pour les individus. Cet effet aléatoire ne sera pas ainsi estimé directement comme c'est le cas pour la composante fixe du modèle (les effets relatifs aux variables  $\beta_{1it} \dots \dots \beta_{Nit}$ , mais plutôt à travers l'estimation de la composante aléatoire du modèle et implicitement sous forme de variance intra individuelle.

C'est à ce niveau qu'on peut introduire la notion du coefficient de corrélation intra class qui, dans le cas de cette forme particulière des modèles hiérarchiques, peut être interprété comme le coefficient de corrélation résiduelle puisque les unités de premier niveau dans notre cas sont les différents points du temps. Si on note par  $\rho$  ce coefficient, on peut l'exprimer comme suit :

$$\rho \equiv Cor(\epsilon_{tij}, \epsilon_{t'ij}) = Cor(y_{tij}, y_{t'ij} \mid x_{1tij} \dots \dots x_{Ntij}; x_{1t'ij} \dots \dots x_{Nt'ij}) = \frac{\tau}{\tau + \theta}$$

Dans ce cas,  $\tau$  mesure la variance de la composante aléatoire de deuxième niveau  $\mu_{ij}$  et  $\theta$  celle de la composante aléatoire globale  $\vartheta_{tij}$ .

Ce coefficient intra class mesure donc la proportion de la variance qui est expliquée par la structure hiérarchisée de la population totale, c'est-à-dire par l'hétérogénéité issue des individus par rapport à la variance globale de la population. Plus particulièrement, elle reflète la corrélation liant deux paires appartenant à la même strate c'est-à-dire pour nos propos à la même période du temps.

Da la première colonne du tableau 8, on peut calculer le coefficient intra class pour notre modèle par  $\rho = \frac{2.639^2}{2.639^2 + 0.941^2} = 0.78$ . Cela signifie que 78 % de la variance globale est potentiellement expliquée par la structure hiérarchisée de cette population, l'hétérogénéité relative aux individus..

L'amplitude de ce coefficient de corrélation indique alors que les paires d'échange sont fortement différenciés par des aspects spécifiques qui ne dépendent pas du temps, et donc non observables au premier niveau, d'où la nécessité de passage à une estimation par strate c'est-à-dire par une approche linéaire mixte.

**Tableau 1 : Estimation sur données de panel de l'équation gravitaire des échanges [1980-2004]**

	<i>Pooled MCO</i>		<i>Between – Paire Effect</i>		<i>Random Effect</i>		<i>Fixed Effect</i>	
	<i>Coef.</i>	<i>p – value</i>	<i>Coef.</i>	<i>p – value</i>	<i>Coef.</i>	<i>p – value</i>	<i>Coef.</i>	<i>p – value</i>
<i>PIB ij</i>	.858	.000	.868	.000	.961	.000	.104	.000
<i>PIB per capita ij</i>	.094	.000	.096	.000	.195	.000	.979	.000
<i>Distance Bilatérale</i>	-1.038	.000	-1.036	.000	-1.019	.000		
<i>Frontière Commune(0 = sinon)</i>	.125	.043	.110	.594	-.529	.012		
<i>Liens Coloniaux(0 = sinon)</i>	1.112	.000	1.105	.000	1.262	.000		
<i>Liens de Langue(0 = sinon)</i>	.521	.000	.526	.000	.436	.000		
<i>Enclavement(0 = sinon)</i>	-.426	.000	-.419	.000	-.545	.000		
<i>Constant</i>	-29.787	.000	-30.223	.000	-21.826	.000		
<i>Adj. R<sup>2</sup></i>	0.751		0.808				0.2773	
<i>No. of cases</i>	54250		54250		54250		54250	
<i>No. of pairs</i>	2170		2170		2170		2170	
<i>F (Chi2 Wald)</i>	23372	.000	1301.56	.000	18353.72	.000	5861.82	.000
<i>rho</i>					.698		.900	
<i>sigma_u</i>					1.151		2.276	
<i>sigma_e</i>					.756		.756	

Source: Les auteurs

Avant d'introduire et de tester les différentes formes de spécificités individuelles, nous procéderons à une estimation de l'équation gravitaire par les MCO avec une disposition de panel pour les données. Comme l'indique la première colonne du tableau 1, la spécification présente une précision statistique pour les différentes corrélations prédites théoriquement liant le volume des échanges aux variables gravitaires.

Le signe de la corrélation reflétant l'effet taille indique que les flux des échanges sont positivement corrélés avec le PIB, mais d'une manière moins que proportionnelle, ce qui corrobore statistiquement la thèse selon laquelle les PVD tendent à s'orienter plus au commerce international. Ces résultats sont de même nature que ceux de Frankel [1997] pour des points distincts du temps dans la période 1965-1992.

Ce constat est confirmé par le signe et l'amplitude du coefficient associé PIB per capita. Le signe positif confirme la tendance des petites économies vers plus de dépendance du commerce international.

Concernant les variables de nature purement géographique, on retrouve les prédictions pour les différentes corrélations liant le volume des échanges aux aspects géographique et de proximité institutionnelle, culturelle et coloniale. Comme facteur d'inertie, les variables approximant l'enclavement physique et la distance géographique affectent considérablement et significativement les échanges. On retrouve ainsi le rôle du transport international comme facteur essentiel pour les expériences d'ouverture internationale des économies.

Concernant les facteurs de similarité institutionnelle tels que les liens coloniaux et culturels jouent un rôle important dans la stimulation des relations commerciales entre les paires de partenaires. On note juste que le coefficient de corrélation relatif à la proximité physique (frontières communes) n'est pas estimé avec précision statistique, la composition particulière de notre échantillon pourrait l'expliquer.

Dans une deuxième étape, on introduit une hétérogénéité spécifique aux individus par des différences dans leurs constantes. La manière la plus facile est de procéder à une estimation de l'équation gravitaire avec variable indicatrice pour chaque paire d'échange. Vu le nombre très élevé de paires dans ce type de régression, les résultats ne sont pas présentés. L'estimation de cette forme de l'équation gravitaire révèle une hétérogénéité forte entre les paires, ce qui est vérifié par les tests individuels des coefficients associés aux variables indicatrices des paires et par le test d'estimation sous contrainte de nullité de la somme de ces coefficients. On qualifie cette méthode de moindres carrés avec variables indicatrices (LSDV).

Une deuxième méthode de tester l'hétérogénéité des paires sera de décomposer l'effet intra et celui inter des variables explicatives du volume des échanges. Cette méthode se limite ici aux variables reflétant l'effet taille et celui du niveau de développement.

Deux spécifications seront donc estimées : la première permet d'estimer l'effet inter de la taille sur le volume des échanges, alors que la deuxième précise l'effet intra des variables. Sur le plan formel, il s'agit de l'équation quantitative (4) représentant l'équation des effets inter paires. Les résultats relatifs à cette forme sont donnés dans la deuxième colonne du tableau 1.

Les résultats sont qualitativement et quantitativement similaires à ceux des estimations par les MCO avec une disposition de panel pour les données de l'analyse. Un effet de taille statistiquement significatif et un effet comme prédit pour les variables d'inertie et de proximité géographique et culturelle.

Pour isoler l'effet intra des variables de taille, il suffit de reprendre l'équation gravitaire avec sa disposition de panel et de centrer toutes les variables par rapport à la moyenne. On retrouve la forme quantitative de l'équation gravitaire 3. Les résultats de cette forme particulière du modèle sont donnés dans la quatrième colonne du tableau 1. Ils indiquent une corrélation significative liant la taille au volume des échanges pour les paires, mais que l'amplitude de cet effet de taille est plus importante, particulièrement pour l'effet indépendant du PIB per capita.

Notons que les indicateurs estimés  $\sigma_u$  et  $\sigma_e$  mesurent dans ce cas respectivement l'erreur de deuxième niveau et celle du premier niveau :  $\hat{\mu}_{ij}$  et  $\hat{\vartheta}_{tij}$ .

D'autre part, l'estimation de cette forme fixe de l'hétérogénéité permet de tester explicitement l'existence d'une hétérogénéité inobservable. En effet, il s'agit d'un test de Fisher qui consiste à tester si toutes les variables indicatrices relatives aux paires d'échange sont égales. Dans notre cas, le test est affirmatif (douzième ligne du tableau 1), et il indique que ces variables indicatrices sont toutes et d'une manière jointe significatif. On déduit alors que les estimations basées sur les MCO ne peuvent pas être sans biais.

Pour relier entre les deux variantes d'effets intra et inter paires, il sera nécessaire de ne pas supposer l'hétérogénéité individuelle en tant que composante fixe du modèle, et d'adopter l'hypothèse alternative que cette hétérogénéité se matérialise par un effet aléatoire. Nous devons alors supposer que cet effet aléatoire des individus est indépendant des autres variables explicatives et de l'erreur de niveau 1. Dans le cas de cette forme, les variables matérialisant l'hétérogénéité des individus sont réintégrées dans l'équation gravitaire quantitative. Cette forme intègre donc les deux types d'effets individuels.

Dans l'équation ci-après, et en suivant Rabe-Hesketh et Skrondal [2005], il est clair que l'estimateur des moindres carrés généralisés constitue une manière de combiner les deux estimateurs.

$$\hat{\beta}_{GLS} = (W + \varphi^2 B)^{-1} (\varphi^2 B \hat{\beta}_B + W \hat{\beta}_W) \text{ Avec } \varphi = \frac{\text{var}(\bar{\vartheta}_{.ij})}{\text{var}(\bar{\vartheta}_{.ij} + \bar{\mu}_{ij})} = \frac{\theta/N}{\theta/N + \tau} = 1 - R$$

$W$  et  $B$ , représentent les matrices des sommes des carrées pour les variables explicatives sous les formes intra et inter pour l'hétérogénéité des individus.

**Tableau 2 : Estimation sur données de panel de l'équation gravitaire des échanges [1980-2004]**

	<i>Random FGLS</i>		<i>Random AR (1)</i>		<i>Endogeneity Analysis</i>	
	<i>Coef.</i>	<i>p – value</i>	<i>Coef.</i>	<i>p – value</i>	<i>Coef.</i>	<i>p – value</i>
<i>PIB ij</i>	.661	.000	.759	.000		
<i>PIB per capita ij</i>	.195	.000	.101	.000		
<i>Distance Bilatérale</i>	-1.01	.000	-1.04	.000	-1.036	0.000
<i>Frontière Commune(0 = sinon)</i>	.529	0.012	.300	0.143	.110	0.594
<i>Liens Coloniaux(0 = sinon)</i>	1.262	.000	1.200	.000	1.105	0.000
<i>Liens de Langue(0 = sinon)</i>	.436	.000	.467	.000	.526	0.000
<i>Enclavement(0 = sinon)</i>	-.545	.000	-.521	.000	-.419	0.000
<i>Constant</i>	-21.826	.000	-24.812	.000	-25.868	0.000
<i>No. of cases</i>	54250		54250		54250	
<i>No. of trade paires</i>	2170		2170		2170	
<i>Chi 2 (Wald) [LR Chi2]</i>	18353.72	0.000	12337.37	0.000	14157.61	0.000
<i>Theta</i>	.869		0.732			
<i>sigma_u</i>	1.151		1.106			
<i>sigma_e</i>	.757		.603			
<i>rho</i>	.697					
<i>rho_ar</i>			.631			
<i>rho_fov</i>			.770			
<i>mn_lgdp</i>					.866	0.000
<i>mn_lgdppc</i>					.096	0.000
<i>dev_lgdp</i>					.104	0.000
<i>dev_lgdppc</i>					.979	0.000

Source: Les auteurs

Les résultats d'estimation de cette forme de l'équation gravitaire sont donnés dans la colonne 1 du tableau 2. Les estimations révèlent une précision statistique pour les corrélations relatives à l'effet de taille, ainsi qu'à l'importance des facteurs géographiques dans la détermination du potentiel d'échange des paires.

De la même méthode d'estimation, on peut dériver le coefficient « $Rho$ » mesurant la part de la variance imputée aux individus et les deux indicateurs « $sigma_u$ » et « $sigma_e$ » indiquant l'amplitude relative des composantes de l'erreur globale en erreur de niveau 1 et une erreur de niveau 2<sup>7</sup>.

<sup>7</sup> Voir Baltagi [2008] pour un exposé théorique riche et des exemples d'application pour les différentes formes des régressions sur données de panel avec effets aléatoires.

**Tableau 3 : Test des effets aléatoires des paires**

<i>Breuschand Pagan Lagrangian multiplier test for random effects</i>		
<i>Ltrade [Paire, t] = Xb + u [Paire] + e [Paire, t]</i>		
	<i>Var</i>	<i>sd = sqrt(Var)</i>
<i>Ltrade</i>	7.674	2.770
<i>e</i>	.571	.756
<i>u</i>	1.325	1.151
<i>Test : Var(u) = 0</i>		
<i>chi2(1)</i>	3.1e + 05	
<i>Prob &gt; chi2</i>	0.000	

Pour tester l'existence des effets individuels, et en procédant à la même méthode d'estimation, Breusch et Pagan [1980] ont proposé un test qui consiste à infirmer l'hypothèse de nullité de la composante de la variance imputable à une hétérogénéité des paires c'est-à-dire :  $\sigma_u^2 = 0$  (voir Baltagi [2004] pp. 59-61 pour un exposé détaillé du test). Le résultat de ce test dans notre cas est donné dans la dernière colonne du tableau 3. Il permet de rejeter à un seuil de risque de 1% l'hypothèse nulle, indiquant l'existence significative des effets aléatoires spécifiques aux individus, et ainsi que l'estimation sur des données de panel est à priori plus appropriée pour l'équation gravitaire des échanges<sup>8</sup>.

Les estimations établies ci-dessus sont basées sur une hypothèse relativement forte, que l'impact intra et inter des variables relatives à l'effet taille dans l'équation gravitaire sont identiques. Plusieurs raisons nous invitent à un être sceptique pour la plausibilité de cette hypothèse. Les deux raisons qui nous semblent essentielles se rapportent aux problèmes de spécification d'une part, et à d'autre liés à l'hétérogénéité inobservable des paires d'échange.

Concernant le premier problème, on note que par nature, l'équation gravitaire fait intervenir des variables qui sont spécifiques aux paires d'échanges, et qui sont donc dans notre cas indépendantes du temps (l'unité de premier niveau).

Ce type de variables affecte et la moyenne des variables relatives à la taille (les variables à deux dimensions) et l'erreur spécifique à une paire particulière  $\mu_{ij}$ . La résultante est ainsi un effet sur la moyenne de la variable dépendante, c'est-à-dire le volume moyen des échanges pour une paire  $ij$ .

---

<sup>8</sup> Une autre estimation sur données de panel a été établie, et elle permet de tester la sensibilité des résultats suite à l'introduction d'un processus auto régressif d'ordre 1 sur les données. Le signe ainsi que l'amplitude des écarts types estimés des variables n'ont pas été affectés.

Ce mécanisme peut expliquer dans notre cadre le fait que l'effet de la taille économique sur le potentiel d'échange d'une paire peut être affecté par les aspects de nature géographique (l'inertie de la distance ou de l'enclavement), institutionnelle (les avantages de proximité coloniale), ou culturelle (les avantages de familiarité de langue).

Le second problème est corollaire au premier. En effet, le mécanisme décrit ci-dessus engendre une différence conséquente de l'effet intra et inter des variables de la taille économique. Ainsi, on aura un effet inter positif de la taille sur le volume des échanges des paires suite à l'omission des variables spécifiques aux paires, et un effet intra négatif spécifique à une paire donnée. Ainsi, le problème lié à l'hétérogénéité inobservée des paires va se matérialiser à travers un problème d'endogénéité des variables de la taille, et ainsi un problème de spécification de l'équation gravitaire.

On propose deux types de tests pour ces deux types de problèmes potentiels. Dans une première étape, on va procéder à une estimation de l'équation gravitaire en intégrant à la fois la moyenne des variables de la taille et l'écart par rapport à la moyenne de ces variables, et ce pour chaque paire. Le modèle à estimer est naturellement celui des effets fixes. Les résultats de cette estimation sont donnés dans la troisième colonne du tableau 2. De ces résultats, on constate que les coefficients respectifs aux variables de la taille en moyenne et en écart par rapport à la moyenne sont sensiblement différents, indiquant ainsi que l'hypothèse d'égalité des effets inter et intra de ces variables est loin d'être plausible. On peut d'autre part tester d'une manière formelle cette hypothèse. Le tableau 4 montre que cette hypothèse est rejetée à un seuil de risque de 1%.

**Tableau 4 : Test d'endogénéité sur les effets intra et inter paires**

$H_0 : mn\_gdp = dev\_gdp$						
<i>Ltrade</i>	<i>Coef.</i>	<i>Std.Dev.</i>	<i>z</i>	<i>p &gt;  z </i>	<i>95% Confidence Int.</i>	
(1)	-.077	0.23	-3.37	.001	-0.123	-0.32
$H_0 : mn\_gdppc = dev\_gdppc$						
<i>Ltrade</i>	<i>Coef.</i>	<i>Std.Dev.</i>	<i>z</i>	<i>p &gt;  z </i>	<i>95% Confidence Int.</i>	
(2)	-.241	0.34	-7.07	.000	-.308	-.174

La seconde étape consiste à suivre le test de Hausman [1978]. En effet, le test de Hausman a fait l'objet d'une diversité d'utilisations et donc d'interprétations. La raison étant que bien que son hypothèse nulle puisse être interprétée avec acuité, l'hypothèse alternative est loin de faire l'unanimité chez ses utilisateurs, puisqu'il s'agit d'un composite d'alternatives.

Et ces alternatives correspondent à plusieurs sources de problèmes économétriques que nous ne discuterons pas ici (cf la thèse de Kamel Ghaddab). Dans ce qui suit, on va essayer de mener une discussion au tour de la problématique relative à la nature appropriée de l'hétérogénéité des paires d'échange dans l'équation gravitaire.

La décision relative à la nature de l'hétérogénéité des paires fixe ou aléatoire pose un grand problème à la fois économétrique et économique. On se réfère pour nos propos sur la discussion développée par Hsiao [2003] et, si on se réfère à Mundlak [1978], l'hypothèse de coefficients aléatoires peut être adoptée dans ce type de modèles.

Dans une dernière étape, on peut recourir à un test de stabilité inter paires des coefficients dans l'équation gravitaire. Le test est connu sous le nom du test de Chow, qui consiste à recourir à une estimation de l'équation gravitaire pour chaque paire d'échange. L'hypothèse nulle à tester est  $H_0: \beta_{ijk} = \beta_k$ . Il s'agit donc d'une statistique de Fisher définie comme le rapport de la somme des carrés des résidus respectives à la régression en panel pour toute la population, et aux différentes paires prises d'une manière isolée.

Pour notre spécification de l'équation gravitaire, la valeur de la statistique calculée de Fisher est de l'ordre de 56.064<sup>9</sup>. Comparée à la valeur critique (2.60), on constate que l'hypothèse nulle de stabilité interindividuelle des coefficients relatifs, aux PIB agrégé et per capita, est rejetée. On conclut alors la nécessité de passer à une forme hiérarchique de l'équation gravitaire où la variable temps va représenter les unités de premier niveau et les paires d'échange comme étant les unités agrégées du deuxième niveau. On dit ainsi que les paires sont emboîtées dans les unités temporelles. Il est crucial de noter la distinction entre cette forme de stratification des individus et celle qui consiste à étudier la structure longitudinale multivariée des données individuelles.

Cette dernière est souvent qualifiée des modèles «Growth Curve», par analogie aux études empiriques médicales visant à étudier l'évolution longitudinale de la croissance du poids des enfants, ou les études spécifiques à l'évolution longitudinale de l'effet des traitements répétés pour les patients<sup>10</sup>.

---

<sup>9</sup> Le test de Chow sur la stabilité inter paires des coefficients a été effectué par un programme sur Eviews 6.

<sup>10</sup> Voir Davis S. [2002] et Frees [2004] pour une synthèse détaillée sur les différentes familles de modèles à coefficients aléatoires appliqués sur des données longitudinales.

En matière économique, le rejet de l'hypothèse d'égalité des coefficients implique que les paires de pays sont différenciées en termes de l'impact de leur taille économique sur le volume des échanges commerciaux.

#### B-4. Estimation de l'équation gravitaire des échanges avec constante et pente aléatoires :

On a procédé à une estimation paire par paire de l'équation gravitaire en vue, d'une part, de la mise en évidence de la variabilité supposée des coefficients de l'équation gravitaire (constante et pentes des variables explicative) et, d'autre part, d'étudier la corrélation entre les constantes spécifiques aux paires d'échange et les pentes respectives aux variables.

Les tableaux 6 et 7 montrent une grande variabilité, à la fois pour les constantes spécifiques aux paires ainsi que pour les pentes relatives à l'effet de taille de l'équation gravitaire. Graphiquement, on utilise généralement un graphique qualifié de «Spaghetti», qui permet de visualiser la dispersion des constantes et pentes dans l'échantillon global des paires. Le graphique 2 est dérivé d'une estimation par les moindres carrés ordinaires de l'équation gravitaire pour chaque paire, et en utilisant la forme classique de l'équation. Ce graphique montre que dans le cas de l'échantillon de référence, et au cours de la période 1980-2004, les paires d'échange sont marquées par une grande hétérogénéité en termes du volume des échanges lorsque les variables explicatives sont nulles (égales à la moyenne), et ainsi de l'effet de taille correspondant.

**Tableau 6 : Statistiques descriptive sur les constantes et pentes des variables (MCO par paire)**

	<i>Obs</i>	<i>Mean</i>	<i>Std.Dev.</i>	<i>Min</i>	<i>Max</i>
<i>Constante</i>	2170	10.41153	2.646708	3.284186	18.90081
<i>Pente GDP</i>	2170	.2047452	5.617666	-33.22283	47.46094
<i>Pente GDP percapita</i>	2170	1.43755	7.482663	-44.40638	43.23457

Pour ce qui est de la corrélation entre les constantes spécifiques et les pentes des coefficients, la figure 3 révèle une dispersion importante pour les paires d'échange dans la période 1980-2004. Numériquement, les tableaux 6 et 7 corroborent cette dispersion, et ce par les variances respectives aux constantes et aux différentes pentes, ainsi que par la covariance entre les constantes et les deux pentes relatives à l'effet de taille. Pour matérialiser la forte dispersion entre les constantes (volume moyen des échanges par paire) et les pentes (produit intérieur brut agrégé et per capita), le graphique 3 montre que cette dispersion est significativement notoire pour le cas de la variable produit intérieur brut des paires.

Figure 2 : Graphique Spaghetti pour l'équation gravitaire

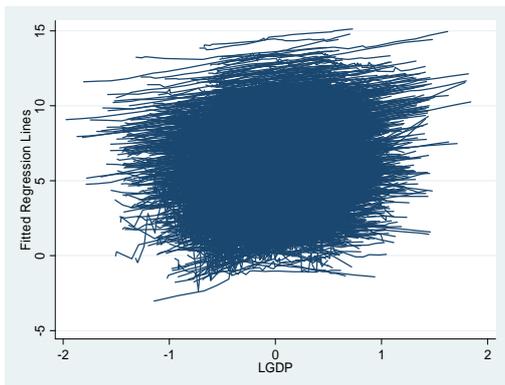


Figure 3 : Pentas vs constantes de l'équation gravitaire

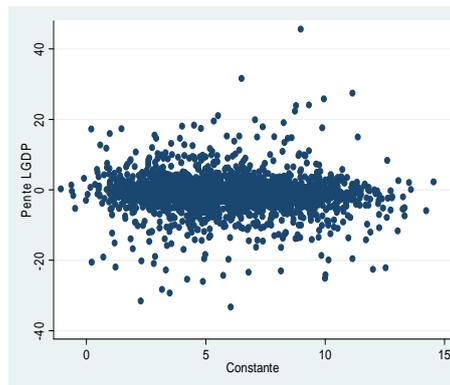


Tableau 7 : Variance et Covariance des constantes et pentes (MCO par paire)

	Constante	Pente GDP	Pente GDP percapita
Constante	7.00506	—	
Pente GDP	-.516645	31.5582	
Pente GDP percapita	.715733	—	55.9903

A ce niveau d'analyse, on va retenir l'hypothèse d'une hétérogénéité aléatoire des paires en termes de constantes et des pentes associées aux variables exogènes. Les hypothèses sous-jacentes aux équations (6) et (7) sont :

- Le volume des échanges bilatéraux des paires de l'échantillon est expliqué par le produit des PIB et PIB per capita. Le signe prédit pour ces deux variables est positif. Ces variables ont été centrées par rapport à la moyenne par paire, pour que les constantes aient une interprétation directe<sup>11</sup>. En effet, la constante représente le volume moyen des échanges pour chaque paire  $ij$ .
- Les variables reflétant les spécificités des paires sont supposées jouer le rôle de variables *modératrices* de la corrélation liant le volume des échanges aux variables de taille. Les paires qui partagent des liens de langue, des frontières communes et une proximité historique et institutionnelle sont supposées avoir une nette interdépendance commerciale, alors que la distance géographique ainsi que l'enclavement territoriale sont supposés freiner les échanges bilatéraux pour une paire  $ij$ .
- En conséquence, les variables gravitaires des échanges modèrent la corrélation des échanges bilatéraux aux variables de la taille,

<sup>11</sup> Le centrage dans le cadre des modèles linéaires mixtes présente une importance capitale, qui peut même conditionner la pertinence des spécifications, résultats et surtout les interprétations. Kreft, Leeuw et Aiken [1994] présentent un exposé détaillé sur les différentes techniques de centrage.

Le tableau 8 résume les estimations de l'équation gravitaire sous ses différentes formes linéaires mixtes<sup>12</sup>. L'estimation de la forme vide du modèle (i.e. sans variable explicative) est donnée dans la première colonne. Elle nous permet de dégager le coefficient intra classe, dont l'amplitude est d'environ 80 %, signifiant que 80 % de la dispersion des paires en termes de leurs échanges commerciaux est expliquée par des facteurs spécifiques à ces paires. Cette amplitude révèle une nature stratifiée des paires d'échange, et elle justifie l'adoption d'une forme hiérarchique de l'équation gravitaire.

**Tableau 8 : Estimation du modèle linéaire mixte de l'équation gravitaire des échanges**

	Empty Model		Random Intercept I		Random Intercept II		Random Coefficients	
	Variable à expliquer : $Ltrade_{ij}$							
Fixed Part	Est	SE	Est	SE	Est	SE	Est	SE
$\beta_{00}$ (_cons)	6.055***	.056	6.055***	.056	15.861***	0.660	15.903***	0.659
$\beta_{10}$ (GDP $ti - GDP .i$ )			.104***	.019	0.104***	0.019	-.524***	0.070
$\beta_{20}$ (GDP per capita $ti - GDP per capita .i$ )			.979***	.030	.979***	.030	1.911***	0.100
<i>Interaction Part</i>								
$\beta_{01}$ (Distance Bilatérale)					-1.174***	0.079	-1.178***	0.079
$\beta_{02}$ (Frontière Commune(0 = sinon))					1.729***	0.427	1.770***	0.427
$\beta_{03}$ (Liens Coloniaux(0 = sinon))					1.883***	0.276	1.838***	0.275
$\beta_{04}$ (Liens de Langue(0 = sinon))					.023	0.137	0.023	0.136
$\beta_{05}$ (Enclavement(0 = sinon))					-1.226***	0.128	-1.233***	0.659
<i>Random Part</i>								
_Cons	2.640	.040	2.641	.040	2.398	0.036	2.400	0.036
(LGDP $ti - LGDP .i$ ), $r_1$							2.635	0.060
(GDP per capita $ti - GDP per capita .i$ ), $r_1$							3.589	0.090
Level 1	.837	.002	.756	.002	.756	0.020	.586	0.001
Log Likelihood	-73315.134		-68028.065		-67819.32		-58481.726	
Deviance	146630.26		136056.13		135638.64		116963.45	
BIC	146663		136110.6		135747.7		117127	
AIC	146636.3		136066.1		135658.6		116993.5	

Source: Les auteurs

Dans la deuxième colonne du tableau 8, est donnée l'estimation de la première forme de l'équation gravitaire avec constante aléatoire et avec deux variables explicatives matérialisant l'effet de taille mesuré par le PIB et le PIB per capita. Le signe des coefficients relatifs aux variables exogènes corrobore le signe prédit théoriquement.

<sup>12</sup> Toutes estimations sur les formes linéaires mixtes ont été établies sur STATA 10.0 à travers les programmes relatifs à la procédure XTMIXED, certains résultats ont été dégagés sur HLM 6.04. Les deux applications donnent les mêmes résultats et on peut même actuellement exécuter HLM sur STATA.

L'amplitude de la variance du premier niveau a diminué, impliquant ainsi que comparativement à la forme vide du modèle, l'intégration de deux variables explicative et l'hypothèse d'une constante aléatoire ont permis d'améliorer l'explication de la dispersion de la variable reflétant les échanges bilatéraux des paires de pays de notre échantillon. La variance estimée des constantes est statistiquement significative au seuil de risque de 1%, permettant ainsi d'infirmer l'hypothèse nulle, et de soutenir plutôt la thèse d'une hétérogénéité aléatoire des paires en termes de leurs volumes moyens des échanges bilatéraux. On peut utiliser l'amplitude relative de la fonction de vraisemblance, qui indique dans ce cas que la première forme de l'équation gravitaire avec constante aléatoire explique mieux la dispersion des échanges des paires que la forme dite vide de cette équation.

Dans une deuxième étape, on propose d'intégrer des variables explicatives au second niveau pour expliquer la dispersion des constantes, c'est-à-dire du volume moyen des échanges par paire. Dans cette forme, on suppose que les coefficients relatifs aux variables explicatives du premier niveau ne sont pas aléatoires, mais plutôt fixes et communs pour toutes les paires de l'échantillon.

**Tableau 9 : Analyse exploratoire des variables explicatives du niveau 2**

<i>Level 1 coefficients</i>	<i>Potential Level 2 Predictors</i>				
	<i>Distance Bilatérale</i>	<i>Frontière Commune</i>	<i>Liens Coloniaux</i>	<i>Liens de Langue</i>	<i>Enclavement</i>
<i>Constante : <math>\beta_{00}</math></i>					
<i>Coefficient</i>	-1.243	4.086	2.105	0.266	-1.076
<i>SE</i>	0.075	0.423	0.284	0.141	0.139
<i>T – Student</i>	-16.471	9.663	7.424	1.879	-7.747
<i>Pente LGDP : <math>\beta_{10}</math></i>					
<i>Coefficient</i>	-0.069	0.726	-0.648	-0.106	-0.064
<i>SE</i>	0.065	0.348	0.231	0.114	0.114
<i>T – Student</i>	-1.073	2.084	-2.802	-0.928	-0.559
<i>PenteLGDP per capita : <math>\beta_{20}</math></i>					
<i>Coefficient</i>	0.185	-0.625	0.435	-0.016	0.139
<i>SE</i>	0.084	0.452	0.300	0.148	0.147
<i>T – Student</i>	2.206	-1.383	1.447	-0.106	0.942

Source: Les auteurs

Toutes les variables autres que celle relative à la langue commune ont des signes de corrélation théoriquement prédits avec une bonne précision. Par opposition à la forme linéaire classique de l'équation gravitaire, cette précision statistique est d'une interprétation particulière. Elle implique en effet que ces variables gravitaires jouent significativement le rôle de variables modératrices pour la corrélation des variables d'échelle sur le volume des échanges.

Un coefficient absolu de la variable distance bilatérale de l'ordre de 1.18, implique que la distance induit une pénalité géographique à l'échange de l'ordre de un à trois. La même proportion de pénalisation est engendrée par l'enclavement géographique des partenaires à l'échange. La proximité de langue, de culture et d'institutions jouent un rôle capital dans la fluidité des échanges bilatéraux des paires de l'échantillon. Force est de noter que l'amplitude des effets des facteurs géographiques, culturels et institutionnels semble sensiblement plus importante que celle dégagée dans les études économétriques visant l'estimation linéaire de l'équation gravitaire.

Dans la quatrième colonne du tableau 8, est estimée une forme modifiée de l'équation gravitaire dans laquelle, on suppose que les constantes des paires et les pentes relatives aux variables de la taille et du niveau du développement économiques sont aléatoires. Quantitativement, il s'agit d'estimer la forme linéaire mixte donnée par les équations 6, 7, 8 et 9. Les résultats d'estimation de cette forme générale de l'équation gravitaire révèlent une précision statistique des corrélations relatives aux variables reflétant l'effet de la taille, celui du niveau de développement, et des variables géographique de proximité et d'inertie du deuxième niveau. Ce constat est basé sur l'infirmité de l'hypothèse nulle pour les variances relatives aux deux pentes.

Les estimations relatives aux deux dernières formes de l'équation gravitaire (constante aléatoire et coefficients aléatoires) ont été basées sur une étude exploratoire, qui permet d'étudier la liste des variables explicatives potentiellement candidats à expliquer les variables endogènes du second niveau, c'est-à-dire les constantes spécifiques aléatoires et les pentes des variables du premier niveau. Sur le plan économétrique, cette analyse exploratoire est basée sur les estimateurs empirique de Bayes des coefficients du premier niveau de la régression. Elle consiste en effet, à étudier la corrélation de ces estimateurs avec les variables spécifiques au second niveau. Dans le cas de notre travail, la méthode proposée par Raudenbush et Bryk [1992] donnée dans le tableau 7 nous révèle que les variables géographiques ont un pouvoir explicatif potentiel statistiquement significatif, particulièrement pour les constantes aléatoires et non pas pour les pentes<sup>13</sup>.

---

<sup>13</sup> Voir paragraphe suivant pour une discussion au tour de l'approche empirique de Bayes dans les modèles linéaires mixtes.

## Conclusion

La qualité globale d'estimation indique une amélioration sensiblement importante, matérialisée par une amélioration de l'indicateur de déviation ( $-2 \text{ Log Likelihood}$ ). En effet le taux de variation de cet indicateur entre le modèle à constante aléatoire et celui à coefficients aléatoires est de l'ordre de 14 %. Cette baisse de la valeur de la fonction de vraisemblance peut être imputée à l'hypothèse permettant de rendre compte de la nature aléatoire des pentes des variables explicatives.

On peut conclure ainsi que l'effet de taille et du niveau de développement sont différenciés d'une manière aléatoire entre les paires d'échange de notre population totale. Dans le même sens, ces résultats révèlent que l'hypothèse dominante dans le cadre des études empiriques de l'équation gravitaire qui défend une symétrie parfaite entre les paires de l'impact du développement économique sur le commerce bilatéral, est loin d'être plausible.

La nature aléatoire des pentes relatives aux variables explicatives est corroborée statistiquement par les variances estimées. En effet, l'amplitude de ces variances dépasse la valeur des erreurs types estimées (2.63 [0.06] pour la variable produit intérieur brut et 3.58 [0.09] pour la variable produit intérieur brut per capita), impliquant ainsi que l'hypothèse de leur nullité est formellement rejetée. La signification statistique de ces variances nous permet de déduire qu'une partie de la dispersion des pentes des paires reste encore inexpiquée.

Pour ce qui est de la qualité comparée des estimations, on constate que la forme générale à coefficients aléatoires semble plus appropriée pour expliquer les échanges bilatéraux des paires dans la période objet d'analyse. Ce constat est basé sur les indicateurs de la fonction de vraisemblance, l'indicateur BIC et celui AIC. Concernant le premier indicateur, il s'agit d'un critère basé sur les valeurs calculées de la fonction de vraisemblance. Il est en effet calculé pour deux formes du modèle : un modèle uniquement avec constante aléatoire, et le modèle avec  $k$  variables explicatives. Numériquement, plus la valeur de cet indicateur est importante, moins le modèle est puissant pour expliquer les données, après avoir retenu l'effet des variables explicatives incluses. Dans notre cas, la spécification qui présente le pouvoir explicatif comparatif le plus important des échanges bilatéraux des paires de l'échantillon de référence est celle relative à l'hypothèse des coefficients aléatoires.

Deux autres indicateurs d'information peuvent être utilisés, notamment le critère d'Akaike (AIC), et le critère d'information de Bayes (BIC). Par opposition au critère de déviation basé sur la valeur de la fonction de vraisemblance, ces deux critères d'information font l'objet d'un usage courant dans le cas des modèles linéaires mixtes, qui sont non emboîtés.

L'emboîtement des modèles est défini dans ce cas par le fait que les modèles doivent intervenir les mêmes variables explicatives dans les deux niveaux de la régression mixte. Ainsi, si la population objet d'analyse est la même, et si la méthode d'estimation des différents modèles est inchangée, ces deux critères restent efficaces pour la sélection des modèles<sup>14</sup>.

Dans notre cas, les amplitudes comparées de ces deux indicateurs nous enseignent que le modèle généralisé à coefficients aléatoires représente la forme qui permet le plus d'expliquer la dispersion globale des échanges bilatéraux des paire dans la période 1980-2004.

## Bibliographie

- Baltagi, B., [2005]. *Econometric Analysis of Panel Data*. Wiley & Sons, Ltd.
- Baltagi, B., [2008]. *Econometrics*. Springer Ed.
- Baum C., [2006]. *An Introduction to Modern Econometrics Using Stata*. Stata Press.
- Bergstrand, J., [1985], «The gravity equation in international trade: some microeconomic foundations and empirical evidence», *Review of Economics and Statistics*, Vol. 67, pp. 471-481.
- Bergstrand, J., [1989], «The generalized gravity equation, monopolistic competition, and factor-proportions theory in international trade», *Review of Economics and Statistics*, Vol. 71, pp. 143-153.
- Bergstrand, J., [1990], «The Heckscher-Ohlin-Samuelson model, the Linder hypothesis, and the determinants of bilateral intra-industry trade», *Economic Journal*, Vol. 100, pp. 1216-1229.
- Breusch, T.S. and A.R. Pagan [1979], «A Simple Test for Heteroskedasticity and Random Coefficient Variation», *Econometrica*, Vol. 47, pp. 1287-1294.
- Bryk, A., & Raudenbush, S. W. [1992]. *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.
- Cohen, J. Cohen, P. West, S. et Aiken, L. [2003]. *Applied Multiple Regression/Correlation Analysis for The Behavioral Sciences*. Lawrence Publishers, London.
- Davis C. [2002]. *Statistical Methods for the Analysis of Repeated Measurements*. Springer Ed.
- Deardorff, A., [1995], «Determinants of bilateral trade: does gravity work in a neoclassical world», *NBER Working Papers*, WP5377, pp. 01-28.
- Demidenko, E. [2004]. *Mixed Models: Theory and Applications*. Wiley Interscience.

---

<sup>14</sup> Les deux critères d'information d'Akaike et de Bayes sont rattachés au critère de la déviation basé sur la fonction de vraisemblance. En effet, si on note par  $d$ , la déviation ( $-2 \text{ Log Likelihood}$ ) :  $AIC = f(d)$  et  $BIC = f(d)$ . Notons juste que pour que ces critères soient fiables pour la comparaison des variantes d'un modèle linéaire mixte qui ont des parties fixe et aléatoire différentes, il faut que l'estimation soit établie par la version complète de maximum de vraisemblance. La raison étant que la version restreinte de cette méthode ne tient compte que de la partie aléatoire du modèle. Voir Hox J. [2002] pour une discussion détaillée.

- Frankel Jeffrey. [1999]. *Regional Trading Blocs in the World Economic System*. Institute of International Economics.
- Frankel, J. et Rose, A., [2002], «An Estimate of the Effect of Common Currencies on Trade and Income», *Quarterly Journal of Economics*, Vol. CXVII, pp. 437-466.
- Frankel, Jeffrey A. et David Romer. [1999], «Does Trade Cause Growth? », *American Economic Review*, Vol. 89, pp. 379-99.
- Frees, E.W. [2004], *Longitudinal and panel data: Analysis and applications in the social sciences*. Cambridge: Cambridge University Press.
- Fujita, M., Krugman, P, and Venables, A.J. [1999], *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press.
- Goennar C. [2006], «A Hierarchical Linear Model of Bilateral Trade: The Role of Geography and Institutions», *Working Paper*, University of North Dakota.
- Harris, M., et Matyas, L. [1998]. «The Econometrics of Gravity Models », Melbourne Institute *Working Papers 5/98*.
- Hausman, J. [1978], «Specification Tests in Econometrics», *Econometrica*, Vol. 46, pp. 1251-1271.
- Hox, J. [2002]. *Multilevel Analysis: Techniques and Applications*. Lawrence Erlbaum Associates, Inc., Publishers.
- Hummels D. et Levinsohn, J. [1995], «Monopolistic competition and international trade: reconsidering the evidence», *Quarterly Journal of Economics*, Vol. 110, pp. 799-836.
- Kreft, I., Leeuw, J. et Aiken, L. [1994], «The effect of different forms of centering in hierarchical linear models», *NISS Technical Report N°30*, National Institute of Statistical Sciences.
- Krugman, P., [1991], *Geography and Trade*, New York, MIT Press.
- Krugman, P., M. Obstfeld, [2003], *Économie internationale*, de Boeck, 4éd,
- Leamer, E., [1974], «The commodity composition of international trade in manufactures: an empirical analysis», *Oxford Economic Papers*, Vol. 26, pp. 350-374.
- Linnemann, H., [1966], *An econometric study of international trade flows*, Amsterdam: North Holland.
- Little, R., Milliken, G., Stroup, W., Wolfinger, R. et Schabenberger. [2006], *SAS for Mixed Models*. SAS Press.
- Nicita A. et Olarreaga [2006], *Trade, Production and Protection*, World Bank.

- Pöyhnen, P., [1963], «A tentative model for the volume of trade between countries», *Weltwirtschaftliches Archiv*, Vol. 90, pp. 93-99.
- Rabe-Hesketh et Everitt Bian. [2007]. *A Handbook of Statistical Analysis Using Stata*. Chapman & Hall Ed.
- Rabe-Hesketh et Skrondal, H. [2007]. *Multilevel and Longitudinal Analysis Using Stata*. Stata Press.
- Raise, S. et Duan, N. [2003]. *Multilevel Modeling: Methodological Advances, Issues and Applications*. Lawrence Publishers, London.
- Redding S. et Venables A. [2004], «Economic Geography and International Inequality», *Journal of International Economics*, Vol. 62, 55-84.
- Redding, S. and Venables, A. [2001], «Economic Geography and International Inequality», *CEP Discussion Paper*, 495, London School of Economics.
- Rose, A., [2000], «One Money, One Market? The Effect of Common Currencies on International Trade», *Economic Policy*.
- Rose, A., [2001], «Currency Unions and Trade: The Effect is Large», *Economic Policy*.
- Rose A., [2002], «Estimating Protectionism Through Residuals from the Gravity Model», *Working Paper*, Haas School of Business.
- Samuelson Paul A., [1954], "The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments," *Economic Journal* 64, (June), pp. 264-289.
- Savage, I. R. et Deutsch, K. W., [1960], «A statistical model of the gross analysis of transaction flows», *Econometrica*, Vol. 28, pp. 551-572.
- Tinbergen, J. [1962], *Shaping the world economy: suggestions for international economic policy*, New York: The Twentieth Century Fund.
- Wei, Shang-Jin. [2001], «Good Governance and Natural Openness», *NBER Working Paper*, W7765.

### **Annexe : Liste des pays de l'échantillon**

Afrique de Sud, Centre Afrique, Albanie, Algérie, Allemagne, Angola, Argentine, Australie, Autriche, Bahreïn, Bangladesh, Belize, Bénin, Botswana, Brésil, Bulgarie, Burkina Faso, Burundi, Cameroun, Canada, Chile, Chine, Chypre, Colombie, Comores, Corée de Sud, Costa Rica, Côte d'Ivoire, Danemark, Djibouti, Equateur, Égypte, El Salvador, Émirat Arabe Unie, Espagne, Fiji, Finlande, France, Gabon, Ghana, Grèce, Guatemala, Guinée, Haïti, Hollande, Honduras, Hong Kong, Hongrie, Îles Maurice, Îles Salomon, Inde, Indonésie, Iran, Irlande, Island, Israël, Italie, Jamaïque, Japon, Jordanie, Kenya, Kuwait, Lao People, Le Royaume Uni, Les Etats Unis d'Amérique, Libéria, Madagascar, Malawi, Malaysia, Mali, Malte, Maroc, Mauritanie, Mexique, Mongole, Mozambique, Népal, Nicaragua, Niger, Nigéria, Norvège, Nouvelle Zélande, Oman, Pakistan, Papua New Guinée, Paraguay, Pérou, Philippine, Pologne, Portugal, République de Congo, République Démocratique de Congo, Rwanda, Saoudite Arabe, Sénégal, Seychelles, Siéra Lionne, Singapour, Soudan, Sri Lanka, Suède, Suisse, Suriname, Syrie, Tanzanie, Tchad, Thaïlande, Togo, Trinité et Tobago, Tunisie, Turquie, Uganda, Vietnam, Zambie et Zimbabwe.