

# Note de Synthèse : Étude sur les déterminants économiques du taux de migration en 2015

MOUNIER TEBAS Teodoro

*Université de Nantes*

Encadré par Mr Darné Olivier

---

## Abstract

Le travail suivant consiste à découvrir quels sont les déterminants économiques responsables des différents taux d'immigration entre les pays du monde. L'étude commence par un résumé basé sur les études antérieures liées à ce sujet. Puis procède à la sélection avec justification des variables nécessaires à une estimation économétrique. Nous avons développé une analyse économétrique complète collectant les informations de 109 pays en 2015 à partir des bases de données de la Banque mondiale et de l'OCDE. Le programme Rstudio a été utilisé pour la manipulation des données et les modélisations économétriques. Sur les neuf variables potentiellement explicatives, cinq d'entre elles ont été retenues. Quatre modèles ont été estimés par les moindres carrés ordinaires, mais trois seulement sont valides. Nous avons fait le choix de conserver la dernière estimation où l'on observe des résultats cohérents avec la théorie. Notre modèle le plus convaincant n'expliquera que 30 % de la variance du taux d'immigration.

*Keywords:* Migration internationale, Analyse économétrique, Modélisation MCO, Rstudio.

---



## 1. Introduction

Tout au long de la vie l'être humain s'est sédentarisé petit à petit mais les migrations n'ont jamais cessé et elles ont évolué au fil du temps. L'Europe a été au XIX<sup>e</sup> siècle la plus grande source d'émigration vers le "nouveau monde". Entre 1800 et 1900, environ 60 millions d'Européens se sont installés sur d'autres continents dans l'espoir d'améliorer leur vie en fuyant les crises endémiques, les persécutions politiques et religieuses ainsi que l'appauvrissement. Aujourd'hui, les continents à l'origine des migrations seraient plutôt l'Afrique et l'Amérique latine, l'Europe et l'Amérique du Nord étant les continents "d'accueil" de cette migration.

Contrairement aux analyses classiques qui tentent d'expliquer les déterminants des flux migratoires, nous avons

essayé d'expliquer le stock de migrants, c'est-à-dire : **pourquoi y a-t-il des pays avec un taux d'immigration plus élevé que d'autres ? Et quels sont les déterminants économiques qui peuvent l'expliquer ?** Notre analyse n'a pas consisté à trouver le meilleur modèle économétrique pouvant expliquer les déterminants des flux migratoires, comme le font les analyses de Moreno Torres et Péridy. Mais plutôt d'étudier la significativité de ces mêmes variables économiques pour expliquer la persistance des migrants dans certains pays plus que dans d'autres. Autrement dit, nous allons voir si les mêmes variables économiques qui expliquent habituellement les flux migratoires sont toujours valables pour expliquer le stock de migrants.

Dans notre étude, nous n'avons pas pu faire de différence entre un migrant et un réfugié car le taux d'immigration, c'est-à-dire le pourcentage d'immigrés et de réfugiés dans la population totale, ne prend pas en compte cette différence, ce qui dans notre analyse est primordial. En effet, la différence réside dans les causes de leur fuite. La confusion est faite car le résultat est le même. Un immigré est un individu qui quitte son pays pour s'installer dans un autre pour des raisons économiques. Le réfugié fuit son pays en raison de situations qui pourraient mettre sa vie en danger, qu'il s'agisse de persécutions politiques, de catastrophes naturelles ou de guerre. De plus, les statistiques peuvent être compliquées lorsqu'un immigrant devient un réfugié pendant la migration. Ou l'inverse, lorsqu'un réfugié ne veut pas retourner dans son pays parce qu'il a trouvé

une meilleure situation économique dans le pays d'accueil.

## 2. Résumé du cadre théorique

### 2.1. Théorie traditionnelle et modèles d'auto-sélection

Les théories traditionnelles défendent que les déterminants de la migration dépendent du différentiel de salaire entre le pays d'origine et de destination, net des coûts de migration. Ces théories ont été développées par Larry A. Sjaastad. Harris et Todaro, avec leur théorie sur le commerce international, ont ajouté que si les pays ne diffèrent que par leurs dotations en facteurs de production (travail et capital), alors le commerce international doit être égal au prix des facteurs, ce qui fait du commerce une alternative aux migrations. **les modèles d'autosélection de Borjas** ont permis de trouver de nouveaux déterminants de la migration, comme la distribution inégale des salaires (gini) ou les qualifications des migrants. Borjas complètera sa théorie de l'auto-sélection par celle de Welfare Magnets en 1999. Une fois que le migrant s'est auto-sélectionné, il choisira le pays dans lequel les prestations sociales sont les plus élevées pour s'y installer. Les théories sur la fuite des cerveaux viendront compléter ces modèles.

### 2.2. Théorie néo-classique et keynésienne

Iván Moreno Torres et Guillem López Casasnovas résument les théories sur la migration d'une autre manière. Ils partent de la **théorie économique néoclassique** qui prône qu'il y aura migration tant que le bénéfice d'émigrer sera supérieur aux coûts, considérant le salaire comme le seul bénéfice et le transport ou quitter la famille comme coût. Les travailleurs des pays où les salaires sont bas, c'est-à-dire où le facteur travail est supérieur au facteur capital, se déplaceront vers des pays où les salaires sont élevés, c'est-à-dire où le facteur travail est inférieur au facteur capital. Un équilibre sera atteint lorsque les écarts de salaires entre pays correspondent au coût de la migration (Massey et al., 1993). **théorie économique keynésienne** ajoute seulement que le travailleur sera attiré par le salaire nominal et non par le salaire réel, ce qui remet en cause l'équilibre néoclassique.

### 2.3. Théorie du marché de travail

La **Théorie du double marché du travail** explique que la migration dépend de la demande de travailleurs des pays industrialisés qui doivent pourvoir les emplois les plus difficiles sur le marché du travail (facteurs "pull"). Et ce ne sont pas tant les faibles salaires ou le chômage dans les pays moins développés qui favorisent cette migration (facteurs "push").

### 2.4. Théorie de la nouvelle économie de la migration

La **nouvelle économie de la migration**, défend que la décision d'émigrer ne vient pas tellement d'une décision individuelle mais plutôt d'une décision conjointe des membres d'un ménage afin de diversifier les risques que

peuvent causer les défaillances du marché et ainsi sécuriser une source d'argent. Un des exemples est l'accès au crédit. Cette décision permet aux familles de réduire leur dépendance à la situation économique d'un seul pays. La décision d'émigrer ne doit pas toujours être prise dans un autre pays mais peut être interne. Cette théorie de Stark et Taylor nous apprend que lorsqu'il n'y a pas d'évaluation du capital humain par le pays d'accueil, l'individu peut choisir de rester dans son pays pour accéder à des emplois qui valorisent sa formation.

### 2.5. Théorie du système mondial

Nous terminerons notre résumé des théories par **La théorie du système-monde**. Cette théorie affirme généralement qu'il y a une migration accrue entre les pays qui ont des relations commerciales en raison de la mondialisation. Plus les deux pays sont proches (distance, ancienne colonie, culture, langue, etc. similaires), plus le nombre de migrants est important.

## 3. Construction de la base

### 3.1. Comment a-t-elle été réalisée ?

Afin d'expliquer les niveaux des taux d'immigration dans les pays observés, nous avons choisi des variables qui ont en théorie un impact sur les migrations au niveau global d'un pays. Certaines de ces variables font référence aux théories traditionnelles où la décision de migrer est prise avec une analyse coûts-avantages. L'avantage étant le niveau de salaire obtenu avec la variable PIB par habitant et les coûts étant le chômage (taux de chômage). A ces variables nous en avons ajouté d'autres comme la balance commerciale et les investissements directs étrangers qui renvoient à l'ouverture du pays avec le monde extérieur. D'autres sont davantage liés aux caractéristiques internes du pays telles que le pourcentage de la population urbaine, la population en âge de travailler ou les dépenses publiques d'éducation. Enfin, nous avons choisi des variables qui collectaient le niveau de pauvreté des individus dans un pays avec l'indice de pauvreté et le coefficient de Gini.

### 3.2. Obtention des données

Nous avons téléchargé pour l'année 2015 les 9 variables qui nous intéressaient directement depuis R avec le package "wbstats" pour un total de 264 observations. Nous sommes passés de 264 observations à 135, en supprimant les pays qui n'allaient pas être pertinents dans l'analyse en raison de leur petite taille et de leur manque d'informations.

Pour le coefficient de Gini, l'indice de pauvreté et les dépenses publiques d'éducation, nous n'avons pas toutes les valeurs pour de nombreuses observations pour 2015, mais la Banque mondiale disposait de ces informations pour les années précédentes. Nous avons choisi de prédire ces valeurs en calculant la moyenne de ces variables pour chaque pays au cours des années précédentes. Une fois

la prédiction faite, nous avons remplacé les valeurs manquantes par ces valeurs prédites. Voyant que les valeurs de l'indice de pauvreté manquaient encore pour des pays importants tels que "Australie" ou "États-Unis", nous avons décidé de compléter la base de données avec les données de l'OCDE. Cela suppose un biais dans l'analyse car il existe deux sources avec des processus différents dans la collecte et le traitement des données. Mais cela a également permis de rajouter des informations dont nous avons besoin pour continuer le travail. Nous avons finalement supprimé les valeurs manquantes restantes et nous sommes passés de 135 à 109 observations.

Variables	Description	Signe	
	Mesure	Espéré	Observé
tasa_migrantes	% de la pob	-	-
pib_pc	\$/habitant	> 0	> 0
pob_desemp	% de la pob act	< 0	< 0
pov	% de la pob	< 0	< 0
gini	0<gini<100%	< 0	< 0
pob_edad_trab	% de la pob	-	> 0
gastopub_educ	% del PIB	> 0	> 0
bc	% del PIB	> 0	> 0
pob_urb	% de la pob	> 0	> 0
ide	% del PIB	> 0	-

TABLE 1: Tableau résumant les variables de la base de données

Une fois les données récoltées nous avons observé les liens qui pouvaient exister entre chacune des variables explicatives avec la variable à expliquer "tasa\_migrantes" (*colonne Observée*), dans le but de vérifier si visuellement la théorie disait vrai (*colonne Espéré*). La table précédente, résume les noms des variables attribuées à chacune d'entre elles avant l'estimation économétrique ainsi que leur unité de mesure.

## 4. Méthodologie économétrique

### 4.1. Un modèle MCO pour étudier ce phénomène

La méthodologie économétrique nous a permis de répondre à notre sujet en transformant notre problématique en quelque chose qui peut se résoudre statistiquement. Dans notre problématique nous sommes face à deux phénomènes que nous souhaitons comprendre : à la fois déterminer si des facteurs économiques peuvent expliquer les différences dans les taux de migrations entre les pays, ainsi que de vérifier la véracité des théories liées à ce sujet.

Le choix de la méthode économétrique utilisé dépend de plusieurs facteurs qui sont, les objectifs de l'analyse ainsi que la structure et le type de variable à étudier. Dans notre cas nous avons choisi la méthodologie des moindres carrés ordinaire. En effet, l'objectif n'était pas de trouver le meilleur modèle prédictif mais plutôt de vérifier l'impact des déterminants économiques dans le taux de migrations afin de mettre en évidence la véracité ou non des théories liées à ce sujet. De plus, nous étions face à des données

transversales<sup>1</sup> (une seule année, 2015) et non temporelle avec une variable à expliquer, continue et non discrète.

### 4.2. Hypothèses liées aux modèles estimés par les moindres carrés ordinaire

L'inconvénient des modèles estimés par la méthode des MCO c'est qu'ils sont soumis à un certain nombre d'hypothèses fortes qu'ils doivent impérativement remplir pour pouvoir en tirer des conclusions. Ces hypothèses sont expliquées ci-dessous.

Le modèle doit être **bien spécifié**, c'est-à-dire qu'il existe une relation linéaire entre les variables, entre les paramètres et qu'il n'y a pas d'omission ou de redondance de variables. Le test Reset Ramsey nous a permis de vérifier si nos modèles étaient bien spécifiés ou non. Il ne doit pas non plus y avoir de **multicolinéarité** parfaite entre les variables exogènes, en d'autres termes, le coefficient de corrélation simple entre deux variables explicatives doit être différent de l'unité. Nous l'avons vérifié avec le facteur d'inflation de la variance, VIF.

Les dernières hypothèses concernent les perturbations aléatoires. Premièrement, les résidus doivent suivre une **distribution normale**. Nous pouvions le vérifier avec le test de Jarque-Bera mais ce test n'est pas nécessaire lorsque l'échantillon est supérieur à 30 observations ce qui était le cas dans notre étude avec 96 observations. Nous devons vérifier se que l'on appelle l'**homoscédasticité** des résidus ce qui signifie que la variance du terme d'erreur doit rester constante pendant toute la période de l'échantillon. Plusieurs tests existent pour cette hypothèse, mais nous avons utilisé ici le test de Breusch-Pagan. Enfin nous avons fini notre série de tests en regardant la distance de Cook qui nous renseigne sur les **observations d'influences**. L'élimination d'un pays ne doit pas, en principe, affecter la précision de la régression. Si toutes ces conditions n'étaient pas vérifiées, le modèle ne pouvait être valide.

### 4.3. Différentes spécifications utilisées pour l'estimation

Après avoir étudié les variables en regardant leur distribution et les corrélations entre elles nous avons commencé les estimations. Au total nous avons réalisé 4 estimations dont seulement trois étaient validées. Nous n'avons pas inclus la population au chômage susceptible d'endogénéité ainsi que le taux de pauvreté qui récupérerait une grande partie de l'information apporter par la variable gini, tout comme la variable ide avec la balance commerciale. Enfin, nous avons retiré le PIB par habitant car il y avait une trop forte corrélation avec l'ensemble des variables. Nous décrivons ci-dessous le choix des spécifications pour chacun des modèles estimés :

- Un modèle MCO1 où l'on a essayé d'expliquer le taux d'immigration par le coefficient de gini, la population

1. On entend par "données transversales" les données relatives à un moment précis ou à une période donnée.

en âge de travailler, les dépenses publiques en éducation, la balance commerciale et enfin la population urbaine. Ce modèle n'a pas passé le test de bonnes spécifications ainsi que d'homogénéité des résidus. Nous avons donc été obligé d'estimer un autre modèle avec une spécifications différentes.

- Un modèle MCO2 où cette fois-ci au lieu d'expliquer le taux d'immigration, nous avons tenté d'expliquer le logarithme du taux d'immigration en conservant les mêmes variables explicatives que MCO1. Ce nouveau modèle a permis de résoudre le problème de la mauvaise spécification ainsi que le problème d'hétéroscédasticité qu'il y avait précédemment. Nous nous sommes rendus compte que la balance commerciale était l'une des variables qui provoquait le plus cette hétéroscédasticité, c'est pourquoi nous avons décidé de réestimer d'autres modèles sans la balance commerciale.
- En ce qui concerne le troisième modèle MCO3, nous avons fait l'hypothèse que toutes les variables explicatives transformées en logarithme pouvaient expliquer le logarithme du taux de migrants par pays. Nous avons réalisé un modèle que l'on nomme **log-log** dans le jargon.
- La dernière modélisation s'est contenté de reprendre MCO3 mais de conserver seulement le logarithme des dépenses publiques en éducation comme variable explicative à transformer. Les autres faisaient l'hypothèse de leur linéarité vis-à-vis du logarithme du taux de migrants. Cette estimation s'est avérée la plus correcte en terme d'explications de la variance du taux de migrants avec une explication à 31.91%.

Modèle	Variables	Coef	Résultat
MCO2	gini	-0.062***	✓
	gastopub_educ	0.210*	✓
	pob_edad_trab	-0.055*	X
	bc	0.01	-
	pob_urb	0.020*	✓
MCO3	log(gini)	-2.549***	✓
	log(gastopub_educ)	0.960**	✓
	log(pob_edad_trab)	-3.468*	X
	log(pob_urb)	1.044**	✓
MCO4	gini	-0.066***	✓
	log(gastopub_educ)	0.880*	✓
	pob_edad_trab	-0.053*	X
	pob_urb	0.023**	✓

TABLE 2: Tableau résumant les principaux résultats des estimations valides

Si nous procédons à l'explication d'un coefficient présent dans la table ci-dessus nous pouvons constater que pour MCO4, le coefficient de gini est significatif à n'importe quel niveau de signification et a également le signe négatif attendu. Le modèle prédit que, pour un pays, lorsque

son coefficient de Gini augmente d'un point de pourcentage, son taux d'immigration diminue de 6,6 %. Le résultat est conforme à la théorie qui supposait une relation décroissante entre le taux d'immigration et une répartition inégale des revenus.

Contrairement au coefficient de gini, la population en âge de travailler est une variable qui nous a donné un coefficient négatif alors que nous attendions une relation positive. En effet, nous avons comme intuition qu'un pays avec un pourcentage d'actifs plus élevé avait un pourcentage plus élevé d'immigrants que les autres. Dans notre modèle, si la population active augmente d'un point de pourcentage, le taux d'immigration baisse de 5,3 % avec un niveau de signification de 10 %. Ce résultat va à l'encontre de la théorie développée.

En général pour tous les modèles estimés, des neuf variables explicatives qui devaient expliquer le taux d'immigration, seulement cinq ont été incluses dans les estimations. Nous avons dû supprimer "*pib\_pc*" et "*pov*" pour la corrélation, mais aussi "*pob\_desemp*" et "*ide*" pour l'endogénéité supposée et les valeurs aberrantes respectivement. Sur les cinq estimations, quatre sont significatives et seules "*gini*", "*gastopub\_educ*" et "*pob\_urb*" sont cohérentes avec ce qui est prédit par la théorie.

## 5. Conclusion

Avec cette analyse, nous avons tenté d'expliquer les différences dans les taux d'immigration entre les pays du monde en 2015. Nous nous sommes concentrés sur les variables économiques pour étudier si les déterminants économiques pouvaient affecter les migrations. Nous avons utilisé la méthode de régression linéaire multiple pour estimer les modèles. Sur les quatre estimations, nous avons trouvé un modèle expliquant environ 30% de la variance du taux d'immigration pour 109 pays. Les variables influençant significativement le taux d'immigration sont l'indice de Gini, la population entre 15 et 64 ans, les dépenses publiques en éducation et enfin la population urbaine.

Mon étude vient compléter d'autres études réalisées sur ce sujet, confirmant l'importance de certaines variables pour expliquer le taux d'immigration. De plus, nous avons vérifié que les théories étaient conformes aux données empiriques de 2015. Sur les quatre variables estimées, trois ont un signe cohérent avec la théorie. La population entre 15 et 64 ans, c'est-à-dire la population active, est la seule variable avec un signe opposé à ce qui était attendu. Nous avons supposé qu'un pays avec une main-d'œuvre importante était un pays en développement, attirant ainsi l'immigration étrangère à la recherche d'un travail. On pourrait expliquer la relation négative de la façon suivante. Les pays que nous considérons en voie de développement en 2015 sont des pays qui tentent de sortir de la pauvreté, et ils ont tendance à avoir une population jeune avec en conséquence un pourcentage élevé de la population active. Mais une population active ne signifie pas nécessairement une population avec des emplois et des opportunités.