

**Economics School of Louvain - ESL  
Economics School of Namur - ESN**

# **Estimation des déterminants de l'immigration clandestine vers l'Europe**

**A partir des données Frontex 2017**

Auteur : Clément PIERRET

Directeur : Frédéric DOCQUIER

Lecteur : Gonzague VANNOORENBERGHE

Année académique 2018-2019

Master en sciences économiques – 120 crédits – Finalité : Spécialisée

## **Avant-Propos**

Avant toute chose, je souhaiterais remercier les personnes ayant contribué, de manière directe ou indirecte, à l'élaboration et à la finalisation de ce mémoire.

Je pense principalement au professeur Frédéric DOCQUIER, mon directeur de mémoire, dont les conseils et les remarques ont su m'orienter tout au long de ce travail. Je le remercie sincèrement pour sa disponibilité et sa compréhension qui m'auront encouragé à aller au bout de ce projet.

Mes remerciements vont également à Gonzague VANNOORENBERGHE qui a accepté de se libérer pour être mon lecteur, et cela dans des délais particulièrement courts.

Merci aussi à Yannik SCHENK, mon assistant de mémoire, pour sa disponibilité et son aide apportée dans le cadre de mes analyses économétriques.

Je tenais également à remercier toute l'équipe administrative de l'ESL pour son écoute et son aide auprès des étudiants. J'en ai bénéficié plusieurs fois et j'ai véritablement apprécié leur contact bienveillant tout au long de mes études.

J'ai aussi une pensée pour tout mon entourage, famille, proches et amis, qui m'a encouragé et soutenu durant le processus de ce mémoire, qui fût par moment une véritable épreuve pour moi. Je pense particulièrement à mon ami Benoit BIHIN, sans qui ce mémoire n'aurait peut-être jamais été conclu.

Enfin, qu'il me soit permis de remercier mes parents, pour m'avoir donné les moyens de poursuivre ces études, et ma mère en particulier, pour toutes ses attentions et son soutien constant.

En espérant que vous apprécierez cette lecture,

Clément PIERRET, août 2019.



## Table des matières

1. Introduction .....	6
2. Données.....	7
2.1. Données Frontex .....	7
2.2. Modèle théorique .....	11
2.3. Les variables explicatives .....	12
3. Analyse empirique.....	13
3.1. Présentation du modèle et de la méthode .....	13
Spécialisation du modèle empirique .....	13
Justification des données.....	13
Le modèle économétrique.....	13
3.2. Le modèle saturé .....	14
3.3 Le modèle non saturé .....	15
3.4. Régressions sur les effets fixes .....	20
3.5. Meilleures spécifications .....	22
6. Conclusion .....	23
7. Bibliographie .....	24

## Table des tableaux

Tableau 1 : Liste des 9 routes migratoire et le total de détections sur chacune .....	8
Tableau 2 : Top 20 des pays d'origine avec les détections les plus fréquentes .....	10
Tableau 3 : Modèle saturé .....	14
Tableau 4 : modèle non-saturé .....	15
Tableau 5 : robustesse à la spécification .....	17
Tableau 6 : spécification de la variable « TDbND » .....	18
Tableau 7: spécification de la variable « intensité conflits » .....	19
Tableau 8 : OLS sur les effets fixes récupérés ( $\alpha$ ot) .....	20
Tableau 9 : OLS avec spécification.....	21
Tableau 10 : meilleures spécifications .....	22



# 1. Introduction

Le but de ce travail, tout en tenant compte de ses limites, sera d'apporter un élément de réponse à la question « Comment tenter d'estimer les déterminants de l'immigration clandestine vers l'Europe (à partir des données Frontex 2017<sup>1</sup>) ? »

Au départ donc de cette question, l'idée d'utiliser une nouvelle base de données produite par Frontex<sup>2</sup>, l'agence européenne de garde-frontières et garde-côtes. Cette organisation est devenue un bras opérationnel de l'UE dans sa tentative de réponse continue aux défis qui se posent à ses frontières extérieures. Au cours de son action, celle-ci a sorti les chiffres du nombre de migrants interpellés alors qu'ils tentaient de rentrer illégalement à l'intérieur de l'espace Schengen. L'utilisation de cette base de données, combinée à d'autres sources de données, pourrait peut-être permettre d'identifier certains déterminants de cette immigration clandestine.

Dans ce travail, nous allons donc tenter d'expliquer plus en détail une des catégories de facteurs qui expliquent les mouvements de l'immigration clandestine. A savoir, la catégorie des facteurs « push » ; ces variables qui poussent des personnes à quitter leur pays afin d'essayer de rentrer en Europe, en dehors des canaux de l'immigration légale. Ces facteurs dépendent des pays d'origine des migrants et ils varient à travers le temps.

Pour mesurer ces facteurs, nous avons dû, d'une part, construire une base de données à partir des informations<sup>3</sup> fournies par l'agence Frontex et la compléter avec d'autres variables jugées pertinentes dans notre visualisation des mécanismes en action, mais aussi jugées compatible avec nos données de base. Et d'autre part, nous avons dû utiliser des modèles statistiques, contrôlant pour certains effets fixes à plusieurs dimensions (principalement le modèle de régression de *ppmlhdfc*<sup>4</sup>) afin de voir quels facteurs peuvent expliquer le nombre variable de migrants sur les différentes routes observées et d'en estimer leurs effets respectifs.

Ce travail s'articule donc en 2 parties. La première partie présente les données<sup>5</sup> fournies par l'agence Frontex, les décrit et les mets en forme. Ensuite, on y présente les modèles théoriques développés pour expliquer les raisonnements suivis dans notre démarche. Et enfin, on y décrit les variables extérieures sélectionnées pour provoquer les effets que l'on veut analyser.

La deuxième partie utilise toutes ces données et ces modèles afin de les faire réagir pour observer et mesurer les effets, attendus ou non, des variables choisies sur notre centre d'intérêt, le nombre de migrants clandestins détectés aux frontières.

---

<sup>1</sup> Frontex data as of 03 November 2017.

<sup>2</sup> Idem

<sup>3</sup> Ibidem

<sup>4</sup> CORREIA S., GUIMARAES P et ZYLKIN TOM. (2019), "PPMLHDFE: Fast Poisson Estimation with High-Dimensional Fixed Effects".

<sup>5</sup> Frontex data as of 03 November 2017.

## 2. Données

### 2.1. Données Frontex

L'idée est donc d'exploiter une base de données récemment produite par Frontex (Frontex data as of 03 November 2017),<sup>6</sup> l'agence européenne de garde-frontières et garde-côtes. En effet, celle-ci a sorti les chiffres du nombre de détections d'entrées illégales à l'intérieur de l'espace Schengen.

Les données sont rapportées à un échelon mensuel (de janvier 2009 à septembre 2017) et par les états membres et les pays associés à l'espace Schengen. Il s'agit du nombre de détections d'entrées illégales à l'intérieur des frontières par des points de passage (BPC, Border Crossing Points) aux frontières extérieures des états membres de l'Union Européenne et des pays associés à l'espace Schengen. Celles-ci sont agrégées pour chaque route migratoire associée à l'endroit où les détections ont été constatées. Pour chaque détection observée, le pays d'origine des individus est également enregistré.

Les données correspondent donc au nombre de détections plutôt qu'à un nombre de personnes. Il est ainsi possible qu'une même personne soit détectée plusieurs fois en rentrant à l'intérieur des frontières à plusieurs reprises. Cependant, il n'existe actuellement pas en Europe de système capable de suivre les mouvements individuels de chaque personne tentant de passer les frontières de manière clandestine. Ainsi, il n'est pas possible d'établir le nombre précis d'individus ayant pénétré illégalement les frontières extérieures. Il faudra tenir compte de cette particularité lorsque les effets seront mesurés dans la partie empirique du travail.

Il a été nécessaire de nettoyer la base de données en enlevant tous les pays pour lesquels le total des détections sur toutes les périodes et toutes les routes ne concernaient pas plus de 50 personnes.

Ont également dû être retirés les pays qui n'étaient pas repris dans la majorité des autres bases de données utilisées, dont les données n'étaient pas complètes ou moins accessibles. Certaines catégories de pays dans la base de données Frontex<sup>7</sup> étaient également difficilement exploitables car spécifiques à des catégories floues de la base de données (ex : « Unspecified sub-Saharan nationals » et « Not specified »).

Après nettoyage, se dégagent les observations des détections constatées sur 9 routes migratoires distinctes et concernant des personnes venant de 103 pays d'origine différents, tout cela s'étalant sur 35 périodes temporelles (les données mensuelles ayant été agrégées en 35 trimestres).

Les routes migratoires sont labellisées en anglais et seront indiquées comme telles pour la suite du travail. Cependant, voici leur traduction<sup>8</sup> ainsi que la description des zones incluses pour chaque route :

- "Eastern Mediterranean" - « la méditerranéenne orientale » : Chypre, la frontière maritime grecque, les frontières terrestres de la Grèce et de la Bulgarie avec la Turquie.

- "Central Mediterranean" – « la méditerranéenne centrale » : les frontières maritimes de l'Italie et de Malte.

---

<sup>6</sup> Frontex data as of 03 November 2017.

<sup>7</sup> Idem.

<sup>8</sup> Ibidem.

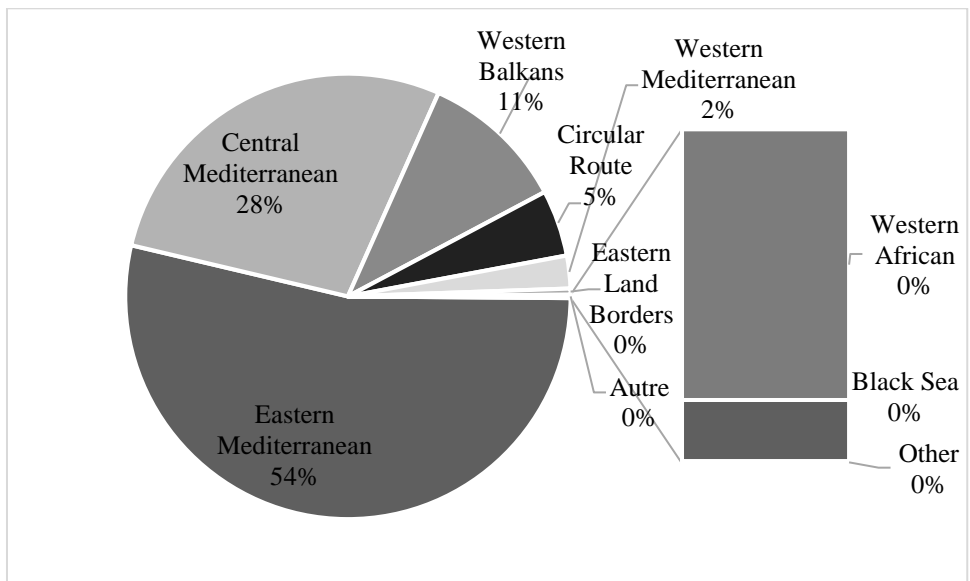
- “Western Balkans” - « (la route) des Balkans occidentaux » : la Grèce, la Bulgarie, la Roumanie, la Hongrie et la Croatie aux frontières terrestres avec des pays de la région des Balkans occidentaux.
- “Circular Route from Albania to Greece” – « la route circulaire de l’Albanie à la Grèce » : la frontière terrestre entre la Grèce et l’Albanie et l’ancienne République yougoslave de Macédoine.
- “Western Mediterranean” - « la méditerranéenne occidentale » : les frontières terrestres et maritimes de l’Espagne, sans les îles Canaries.
- “Eastern Land Borders” - « les frontières terrestres orientales » : la Roumanie, la Hongrie, la Slovaquie, la Pologne, la Lituanie, la Lettonie, l’Estonie, la Finlande et la Norvège, les frontières terrestres avec la Moldavie, l’Ukraine, la Biélorussie et la Fédération de Russie.
- “Western African” – « l’africaine occidentale » : les îles Canaries.
- “Black Sea” : - « la mer Noire » : la Bulgarie et les frontières maritimes de la Roumanie.
- “Other” - « autres » : les zones non comprises dans les autres itinéraires. Par exemple : la mer Baltique, la mer du Nord.

*Tableau 1 : Liste des 9 routes migratoire et le total de détections sur chacune*

Routes migratoires	Détections de 2009 à 2017	% du total
Eastern Mediterranean	1360854	53,57
Central Mediterranean	710795	27,98
Western Balkans	268302	10,56
Circular Route from Albania to Greece	122502	4,82
Western Mediterranean	60410	2,38
Eastern Land Borders	11254	0,44
Western African	4967	0,2
Black Sea	1123	0,04
Other	42	0
Détections totales de 2009 à 2017, tous pays confondus :		2540249

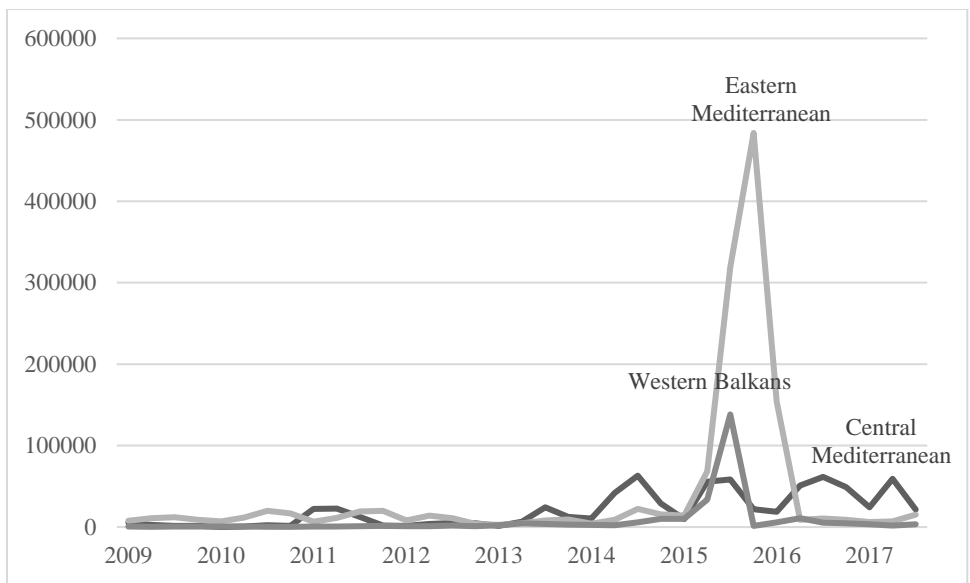
Source : construction par l’auteur.

Figure 1 : la part de chaque route migratoire dans le total des détections



Source : construction par l'auteur.

Figure 2 : Variations temporelles des détections pour le top 3 des routes migratoires



Source : construction par l'auteur.

On voit que la répartition du nombre de détections entre les routes est assez inégale, et qu'un pic de fréquentation se démarque entre 2015 et 2016. Cette hausse du nombre de passages de migrants est flagrante sur la route Eastern Mediterranean (la plus fréquentée en termes absolus) mais elle est également observable sur la route Western Balkans. On note aussi une augmentation des détections sur la route Central Mediterranean à ce moment-là, mais cette route semble aussi réagir à une sorte de cycle de migration car on peut observer une hausse de la fréquentation en été et une diminution en hiver, visible graphiquement depuis 2013

A ces 9 routes de migration, s'ajoute une deuxième variable : le pays d'origine des migrants. Le tableau 1.1 nous donne une idée du poids des 20 nationalités les plus observées chez les personnes tentant de passer les frontières clandestinement. La présence d'individus syriens et afghans est clairement marquée par rapport aux autres origines. A noter que c'est l'Iraq qui vient compléter le top 3, uniquement composés de pays ayant subi des conflits armés violents au cours de la période observée.

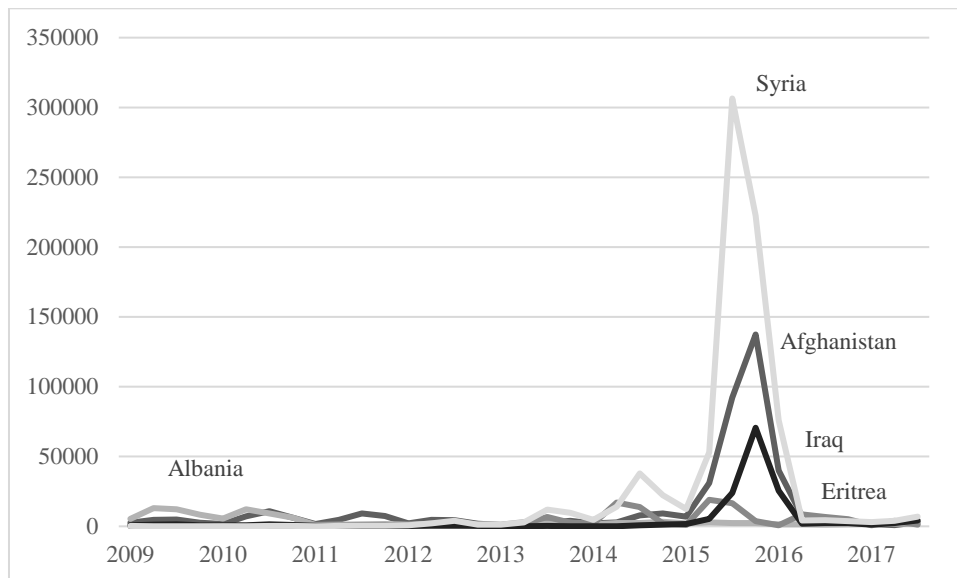
Tableau 2 : Top 20 des pays d'origine avec les détections les plus fréquentes

Pays	Détections de 2009 à 2017	% du total	Pays	Détections de 2009 à 2017	% du total
Syria	811800	31,96	Morocco	42578	1,68
Afghanistan	435593	17,15	Gambia	42072	1,66
Iraq	153846	6,06	Algeria	41783	1,64
Albania	121735	4,79	Mali	40877	1,61
Eritrea	121200	4,77	Palestine	37943	1,49
Pakistan	104402	4,11	Guinea	37345	1,47
Nigeria	100817	3,97	Côte d'Ivoire	36766	1,45
Somalia	63734	2,51	Iran	36433	1,43
Bangladesh	49772	1,96	Sudan	30671	1,21
Tunisia	42866	1,69	Senegal	30527	1,20

Détections totales de 2009 à 2017, tous pays confondus : 2540249

Source : construction par l'auteur.

Figure 3 : Variations temporelles des détections pour le top 5 des pays d'origine



Comme sur les routes migratoires importantes, on observe un pic du nombre de détections entre 2015 et 2016 pour les nationalités les plus représentées (Syrie, Afghanistan, Iraq).

Toutes ces observations reprises dans la base de données Frontex<sup>9</sup> de 2017 présentent un intérêt certain par leur actualité, leur précision et leur nombre. Aussi, pour renforcer la puissance statistique de la base de données, elle sera équilibrée en multipliant chaque composante d'observation par les autres afin d'obtenir toutes les combinaisons possibles de routes, pays d'origine et période de temps (trimestre). La base de données ainsi balancée compte désormais :

9 routes \* 103 pays \* 35 trimestres = 32 445 observations

Chaque observation originellement présente dans la base de données Frontex<sup>10</sup> y trouve sa place, le reste des données est donc composé des 0 générés par l'absence d'observation.

## 2.2. Modèle théorique

La variable que l'on tente d'expliquer est donc le nombre de détections de migrants clandestins, et elle dépend de 3 facteurs : la route empruntée, le pays d'origine des individus qui migrent et enfin la période de temps que l'on observe. Ces 3 facteurs déterminants peuvent être modélisés sous forme de 3 paramètres d'interactions, soit 3 effets fixes à 2 dimensions qui expliquent le nombre de détections comme ceci :  $Y_{ijt} = \alpha_{ij} + \alpha_{it} + \alpha_{jt} + \varepsilon_{ijt}$

Le but de ce travail sera d'explorer le 3<sup>ème</sup> de ces effets fixes et de décortiquer ce  $\alpha_{jt}$  en plusieurs variables explicatives afin d'en estimer les coefficients. Le modèle avec l'effet fixe  $\alpha_{jt}$  ouvert peut se réécrire sous cette forme :

<sup>9</sup> Frontex data as of 03 November 2017.

<sup>10</sup> Idem.

$$Y_{ijt} = \alpha_{ij} + \alpha_{it} + (\beta_1 X1_{jt} + \beta_2 X2_{jt} + \dots) + \varepsilon_{ijt}$$

Et pour inscrire le modèle dans le concret de notre base de données, on fait correspondre la dimension  $i$  à la route migratoire  $r$ , la dimension  $j$  au pays d'origine  $o$  et la dimension  $t$  à la période, soit le trimestre  $t$ , donnant ainsi l'équation :  $Y_{ort} = \alpha_{or} + \alpha_{rt} + \alpha_{ot}$

Traduite en des termes qui ont déjà été énoncés cela donne :

Nbre de Détections = effet fixe  $or$  + effet fixe  $rt$  + variables explicatives  $X_{ot}$

### 2.3. Les variables explicatives

Il s'agit à présent de trouver des variables qui expliquent au mieux cet effet fixe  $\alpha_{ot}$ . Ce sont donc des variables qui peuvent différer d'un pays à l'autre et qui évoluent au cours du temps. Et afin de couvrir au maximum l'effet origine-temps, 5 thèmes, 5 types de variables différents ont été choisis et récoltés pour réaliser notre analyse empirique. Il s'agit du PIB (1), de la population âgée entre 15 et 64 ans (2), de la mortalité infantile (3), du nombre de morts causé par des désastres naturels (4) et enfin de l'intensité des conflits armés dans le pays (5). Voici chaque variable plus en détail :

(1) Le PIB utilisé est le PIB par habitant mesuré en Parité de Pouvoir d'Achat (PPP) et exprimé en dollar international (source : ILOSTAT database<sup>11</sup>).

(2) La population entre 15 et 64 ans est simplement le nombre d'habitants résidants dans un pays et entrants dans cette tranche d'âge. Elle est ainsi choisie car elle correspond à la population en âge de travailler (source : United Nations Population Division<sup>12</sup>).

(3) La mortalité infantile est le nombre de nouveaux nés décédés avant l'âge de 1 an. Elle est exprimée en /1000, c'est donc le nombre de décès par 1000 enfants ayant moins d'1 an (source : UN Inter-agency Group for Child Mortality Estimation<sup>13</sup>).

(4) La variable suivante est intitulée comme telle dans nos analyses : TDbND. Cet acronyme vient de l'expression anglaise « Total Deaths by Natural Disasters », soit le nombre total de morts par désastres naturels. Cette variable peut être décomposée en différents types de désastres naturels qui seront expliqués plus tard (sources : EM-DAT : The Emergency Events Database<sup>14</sup>).

(5) L'intensité des conflits armés est une variable qui peut prendre 3 valeurs sur une échelle d'intensité (0, 1 ou 2) correspondant au nombre de victimes causées par au moins un conflit armé dans le pays. 0 signifie qu'il n'y a pas plus de 25 victimes, 1 signifie que le nombre de victimes est compris entre 26 et 999 et 2 signifie que le conflit a fait au moins 1000 victimes au cours de la période (source : UCDP/PRIO Armed Conflict Dataset<sup>15</sup>).

<sup>11</sup>International Labour Organization, ILOSTAT database. Data retrieved in November 2017.

<sup>12</sup> United Nations Population Division. World Population Prospects: 2017 Revision.

<sup>13</sup> UN Inter-agency Group for Child Mortality Estimation (UNICEF, WHO, World Bank, UN DESA Population Division).

<sup>14</sup> EM-DAT: The Emergency Events Database - Université catholique de Louvain (UCLouvain) - CRED.

<sup>15</sup> Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand (2002) Armed Conflict 1946-2001: A New Dataset. Journal of Peace Research 39(5).

### 3. Analyse empirique

#### 3.1. Présentation du modèle et de la méthode

##### *Spécialisation du modèle empirique*

Le modèle utilisera la méthode du panel selon les différents pays et les différentes routes empruntées en fonction du temps.

Afin d'estimer les coefficients des variables explicatives et afin de contrôler la présence d'effets fixes à plusieurs dimensions on utilisera principalement des régressions de Poisson par maximum de vraisemblance, notamment une forme particulière adaptée aux effets fixes à plusieurs dimensions : la fonction *ppmlhdfe*<sup>16</sup>.

Certaines variables seront transformées pour expliquer au mieux les détections de migrants. On utilise notamment la transformation logarithmique pour les variables concernant le PIB, la population, le TDbND (Total Deaths by Natural Disasters), le TAbND (Total Affected by Natural Disasters), le TDbMétéo (Total Deaths by Meteorological Disasters), le TDbDestructifs (Total Deaths by Destructive Disasters) et le TDbBio (Total Deaths by Biological Disasters). Ces variables et certaines autres déclinaisons seront expliquées plus tard.

##### *Justification des données*

L'essentiel des données utilisées afin de construire notre base de données viennent de l'agence Frontex<sup>17</sup>. Une autre source provient du logiciel EM-DAT<sup>18</sup>: The Emergency Events Database - Université catholique de Louvain (UCLouvain) – CRED, (via le site D. Guha-Sapir - [www.emdat.be](http://www.emdat.be), Brussels, Belgium). Ce site donne accès à la base de données The international disaster Database qui répertorie les effets essentiels de plus de 22 000 catastrophes de masse dans le monde de 1900 à nos jours. La base de données est compilée à partir de diverses sources, y compris des agences de l'ONU, des organisations non gouvernementales, des compagnies d'assurance, des instituts de recherche et des agences de presse.

##### *Le modèle économétrique*

Notre question de recherche est d'évaluer les déterminants pertinents de l'immigration clandestine vers l'Europe, et en particulier ceux qui dépendent des caractéristiques du pays d'origine des migrants et qui varient dans le temps. Ce sont les « push-factors » qui poussent les gens à partir de chez eux afin de tenter de rentrer clandestinement à l'intérieur des frontières européennes. Ces « push-factors » sont les  $\alpha_{o,t}$  du modèle théorique. Mais quel est l'impact relatif de ces facteurs par rapport aux autres effets fixes mentionnés précédemment ? Car si ceux-ci n'expliquent qu'une faible partie du modèle, alors chercher à mesurer leur influence n'a pas grand intérêt.

Pour ce faire, on a d'abord analysé les données selon un modèle saturé qui regroupe tous les

---

<sup>16</sup> CORREIA S., GUIMARAES P et ZYLKIN TOM. (2019), "PPMLHDFE: Fast Poisson Estimation with High-Dimensional Fixed Effects".

<sup>17</sup> Frontex data as of 03 November 2017.

<sup>18</sup> EM-DAT: The Emergency Events Database - Université catholique de Louvain (UCLouvain) – CRED.

effets expliquant notre variable dépendante  $Y_{ijt}$ .

On a estimé les différentes variations des variables avec des effets fixes monotones (1 dimension) sur trois variables : la variable  $o$  qui indique le pays d'origine de l'immigré, la variable  $r$  qui indique la route d'immigration parcourue par l'immigré et la variable  $t$  qui est la variable temporelle.

Ce même tableau a été complété par une analyse à deux dimensions reprenant ces mêmes paramètres. L'effet fixe  $\alpha_{or}$  reprenant donc les variables liées au pays d'origine et de la route empreinte par le migrant. On a effectué le même procédé pour les effets  $\alpha_{rt}$  et  $\alpha_{ot}$ .

### 3.2. Le modèle saturé

Tableau 3 : Modèle saturé

Effets fixes	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A une dimension							
$\alpha_o$	Oui	-	-	-	-	-	-
$\alpha_r$	-	Oui	-	-	-	-	-
$\alpha_t$	-	-	Oui	-	-	-	-
A deux dimensions							
$\alpha_{or}$	-	-	-	Oui	-	-	Oui
$\alpha_{rt}$	-	-	-	-	Oui	-	Oui
$\alpha_{ot}$	-	-	-	-	-	Oui	Oui
Pseudo R <sup>2</sup>	0.4689	0.4825	0.3790	0.7731	0.6461	0.5982	0.9835

Source : construction par l'auteur.

On peut observer sur ce tableau du modèle saturé à effets fixes monotones et à deux dimension un R<sup>2</sup> hautement significatif pour les trois valeurs selon le modèle à deux dimensions avec une valeurs explicative de 0,9855 pour l'ensemble des trois effets fixes.

Ce tableau nous donne une bonne direction pour le reste du travail car il nous permet d'affirmer que les trois variables jouent un rôle significatif dans les détections de migrants de respectivement 0,4689 pour l'effet sur l'origine du pays, de 0.4825 pour la route empreintes et 0.3790 selon le moment de la détection. Les résultats sont sensiblement supérieurs selon le modèle des effets fixes à deux dimensions avec respectivement 0.7731 pour  $\alpha_{or}$  comprenant le pays d'origine ainsi que la route empruntée, 0.6461 pour  $\alpha_{rt}$  comprenant la route et la période de détection et 0.5982 pour  $\alpha_{ot}$  comprenant le pays d'origine et le moment de la détection. Si l'effet fixes  $\alpha_{ot}$  explique à lui seul presque 60% de l'effet total, cela justifie bien un focus sur les facteurs « push » que sont ces  $ot$ .

Même si les trois dernières valeurs sont intéressantes afin d'expliquer l'immigration clandestine vers l'Europe et peuvent s'avérer de bons déterminants au sens statistique, la suite

de ce travail se concentrera sur un seul de trois : les effets fixes  $\alpha_{ot}$  reprenant les effets du pays d'origine et du moment de la détection du migrant. Il peut cependant être intéressant d'analyser les autres facteurs, d'autant que leur pouvoir explicatif est supérieur à celui de notre modèle d'intérêt avec un  $R^2$  de 0.7731 et de 0.6461. Mais, ne disposant pas des ressources nécessaires pour un travail de cette envergure, il a fallu restreindre le travail au seul  $\alpha_{ot}$  qui est également significatif avec un coefficient de détermination de 0.5982. Les deux autres effets fixes pouvant faire l'objet de futures études.

Une fois les effets fixes évalués, l'étude des déterminants de l'immigration clandestine peut s'appuyer sur un modèle non-saturé.

Ce modèle représente une analyse plus fine du modèle précédent, mais avec comme seul effet fixe exploré celui de  $\alpha_{ot}$ , les 2 autres étant contrôlés afin de ne pas polluer les coefficients spécifiques à l'effet origine-temps. On éclate alors l'effet fixe en cinq variables différentes qui tentent de capturer l'essentiel du pouvoir explicatif de l'effet fixe décomposé.

### 3.3 Le modèle non saturé

Tableau 4 : modèle non-saturé

	Benchmark	Sensibilité au retard			
	(1)	(2)	(3)	(4)	(5)
Log (PIB)	-1.449*** (0.005)	-1.272** (0.025)	-0.577 (0.354)	-1.508** (0.015)	-0.807 (0.164)
Log (pop 15-64 ans)	0.928 (0.408)	0.977 (0.393)	0.033 (0.979)	2.648** (0.039)	1.555 (0.198)
Mortalité infantile	-0.026 (0.363)	-0.015 (0.645)	-0.025 (0.435)	0.002 (0.947)	-0.003 (0.921)
Log (TDbND)	0.063*** (0.001)	-0.040** (0.045)	-0.022 (0.186)	0.011 (0.542)	0.011 (0.482)
Intensité conflits	0.672*** (0.000)	0.626*** (0.000)	0.683*** (0.000)	0.644*** (0.000)	0.652*** (0.000)
$\alpha_{or}$	Oui	Oui	Oui	Oui	Oui
$\alpha_{rt}$	Oui	Oui	Oui	Oui	Oui
Pseudo $R^2$	0.9387	0.9371	0.9369	0.9368	0.9360

Entre ( ) se trouvent les  $P > |z|$  valeurs. 17148 observations. Source : construction par l'auteur.

La première colonne est celle des nouvelles variables explicatives. Elle contient les observations concernant les pays de notre base de données, certaines mises en logarithme afin d'en atténuer les extrêmes. La variable population est également mise en logarithme pour les mêmes raisons. La variable de la population entre 15 et 64 ans, c'est-à-dire la population en âge de travailler, a été préférée à la variable de la population totale pour deux raisons principales. D'une part, la majorité des migrants détectés selon l'agence Frontex répond à cette

description. D'autre part, la population hors de ce critère n'est pas déterminante pour cette étude. La mortalité infantile est un indicateur pour estimer les conditions de vie d'un pays liées au niveau de santé publique. Il est donc incorporé au modèle et est un des « push » facteurs qui peut inciter la population à migrer hors de son pays. On apportera une attention particulière à toutes les variables potentiellement significatives d'un départ. La variable « TDbND », *Total Deaths by Natural Disasters*, quantifie le nombre de morts provoqués par une catastrophe naturelle<sup>19</sup>. La dernière variable explicative considérée, l'intensité des conflits se calcule sur une échelle établie par l'agence UCDP GED<sup>20</sup> et dont les données ont été récoltées auprès des trois organismes repris dans la bibliographie en fin de ce travail. La deuxième colonne est une synthèse du modèle non-saturé pour la période t. Elle sera régulièrement reprise pour comparer les différents modèles utilisés. La troisième colonne, ainsi que les suivantes, contiennent les résultats du modèle, mais pour des périodes précédentes à l'exercice t dont les résultats sont observables sur la deuxième colonne. Elles seront utilisées afin de déceler un décalage dans le temps des variables explicatives, soit un éventuel effet à retardement. Pour plus de compréhension, chaque en-tête de colonne de résultat est accompagné d'un chiffre (1 signifie que les résultats sont obtenus sur la période t, 2 signifie que les résultats sont obtenus sur la période t-1 et ainsi de suite).

Les résultats tendent à montrer que le retard n'a pas énormément d'effet sur les coefficients des variables explicatives : ils sont globalement moins forts et très souvent moins significatifs. La seule variable qui reste stable avec un retard de 1, 2, 3 et même 4 périodes, c'est l'intensité des conflits qui garde sa significativité et un  $\beta$  quasi constant. Cela encourage en tout cas à garder le modèle contemporain comme référence car son pseudo  $R^2$  est le meilleur est les effets du PIB sont les plus significatifs (c'est vrai aussi pour l'intensité des conflits, mais cela pour toutes les régressions du tableau 4).

---

<sup>19</sup> La liste des catastrophes naturelles considérées et les données proviennent de la même source, le centre de recherche EM-DAT: The Emergency Events Database.

<sup>20</sup> Högbladh Stina, 2019, "UCDP GED Codebook version 19.1", Department of Peace and Conflict Research, Uppsala University.

Tableau 5 : robustesse à la spécification

	Benchmark	Variables <sup>2</sup>		Variables d'interaction	
	(1)	(2)	(3)	(4)	(5)
Log (PIB)	-1.449*** (0.005)	-6.208 (0.229)	-2.317*** (0.000)	-1.453*** (0.004)	-1.147** (0.028)
[Log (PIB)] <sup>2</sup>	-	0.295 (0.339)	-	-	-
Log (pop 15-64 ans)	0.928 (0.408)	1.132 (0.323)	0.997 (0.402)	0.946 (0.398)	0.935 (0.390)
Mortalité infantile	-0.026 (0.363)	-0.042 (0.214)	-0.255*** (0.000)	-0.026 (0.364)	-0.040 (0.166)
[Mortalité infantile] <sup>2</sup>	-	-	0.001*** (0.000)	-	-
Log (TDbND)	0.063*** (0.001)	0.063*** (0.001)	0.062*** (0.001)	0.065*** (0.001)	0.066*** (0.000)
Intensité conflits	0.672*** (0.000)	0.662*** (0.000)	0.633*** (0.000)	0.674*** (0.000)	1.015*** (0.000)
TDbND*pop 15-64 ans	-	-	-	-8.76e-13 (0.202)	-
Intens.conflits*pop 15-64	-	-	-	-	-1.00e-08*** (0.000)
$\alpha$ or	Oui	Oui	Oui	Oui	Oui
$\alpha$ rt	Oui	Oui	Oui	Oui	Oui
Pseudo R <sup>2</sup>	0.9387	0.9387	0.9395	0.9387	0.9398

Entre ( ) se trouvent les  $P > |z|$  valeurs. 17148 observations. Source : construction par l'auteur.

Dans ce tableau 5, quatre spécifications sont ajoutées une par une afin de voir comment le modèle réagit. D'abord en ajoutant la forme quadratique du log du PIB et celle de la mortalité infantile, puis en ajoutant 2 variables d'interaction. Celles-ci lient la population des 15-64 ans au nombre de morts causés par des désastres naturels et lient cette même population de 15-64 ans à l'intensité des conflits. La seule des ces 4 spécifications qui pourrait être robuste est celle de la mortalité infantile au carré car ses valeurs montrent une bonne significativité. Aussi, le signe négatif de la variable en niveau et le signe positif de la variable au carré peuvent illustrer un effet convexe de la mortalité infantile. Ainsi, une augmentation de la mortalité réduirait les détections jusqu'à un taux de mortalité de 127,5/1000 qui donnerait un effet positif passé ce seuil. Cette idée n'est malheureusement pas très réaliste car les taux de mortalité les plus élevés de la base de données étaient ceux du Sierra Leone avec un maximum de 116/1000 en 2009. Ceux-ci ayant baissé depuis avec un taux maximum de 86/1000 en 2017 pour la Centrafrique.

Tableau 6 : spécification de la variable « TDbND »

	Benchmark (1)	(2)	(3)	(4)	(5)	(6)
Log (PIB)	-1.449*** (0.005)	-1.564*** (0.002)	-1.447*** (0.005)	-1.525*** (0.003)	-1.569*** (0.002)	-1.670*** (0.001)
Log (pop 15-64 ans)	0.928 (0.408)	0.929 (0.399)	1.007 (0.391)	1.055 (0.337)	1.187 (0.309)	1.248 (0.251)
Mortalité infantile	-0.026 (0.363)	-0.028 (0.326)	-0.022 (0.483)	-0.027 (0.325)	-0.023 (0.455)	-0.030 (0.280)
Log (TDbND)	0.063*** (0.001)	-	-	-	-	-
Log (TAbND)	-	0.030*** (0.000)	-	-	-	-
Log (TDbMétéo)	-	-	0.009 (0.735)	-	-	0.015 (0.582)
Log (TDbDestructifs)	-	-	-	0.107*** (0.000)	-	0.107*** (0.000)
Log (TDbBio)	-	-	-	-	-0.070*** (0.008)	-0.067** (0.010)
Intensité conflits	0.672*** (0.000)	0.670*** (0.000)	0.674*** (0.000)	0.680*** (0.000)	0.670*** (0.000)	0.673*** (0.000)
$\alpha$ or	Oui	Oui	Oui	Oui	Oui	Oui
$\alpha$ rt	Oui	Oui	Oui	Oui	Oui	Oui
Pseudo R <sup>2</sup>	0.9387	0.9392	0.9377	0.9398	0.9380	0.9402

Entre ( ) se trouvent les  $P > |z|$  valeurs. 17 148 observations. Source : construction par l'auteur.

Le tableau 6 développe la variable TDbND qui, pour rappel, donne le nombre de morts causés par des désastres naturels (c'est le logarithme de TDbND qui est repris dans les tableaux). On compare ici ses effets avec ceux de sa variante TAbND, ou bien en la décomposant en différents types de désastres naturels. Voici les définitions de ces nouvelles variables :

-TAbND (Total Affected by Natural Disasters) : il s'agit du nombre total de personnes affectées (donc pas mortes) par les mêmes désastres naturels que la variable TDbND.

-TDbMétéo (Total Deaths by Meteorological Disasters) : le nombre de personnes décédées à cause de désastres de type météorologiques (ex : sécheresses, températures extrêmes, feux de forêt, etc.)

-TDbDestructifs (Total Deaths by Destructive Disasters) : le nombre de personnes décédées à cause de désastres causant des dégâts matériels importants (ex : inondations, tempêtes, tremblements de terre, etc.)

-TDbBio (Total Deaths by Biological Disasters) : le nombre de personnes décédées à cause de désastres biologiques (ex : épidémies, attaques animales, infestations, etc.)

Toutes ces nouvelles variables sont prises en logarithme, comme la variable TDbND l'est

aussi. Si l'on compare TDbND et TAbND, ils sont tous deux significatifs mais le coefficient du second est moins fort que celui de notre régression standard. Mais le modèle avec TAbND offre un R<sup>2</sup> légèrement meilleur. Cependant le R<sup>2</sup> est encore plus élevé (le maximum de cette spécification) lorsque l'on décompose TDbND en 3 sous-types de désastres. L'effet de TDbMétéo n'est pas significatif mais TDbDestructifs à lui tout seul est plus fort que TDbND et tout aussi significatif. Quant à TDbBio il surprend par un effet négatif et significatif, exposant peut-être des dynamiques différentes selon le type de désastre naturel. La régression (6) du tableau 6 n'est donc pas sans intérêt.

Tableau 7: spécification de la variable « intensité conflits »

	Benchmark	Variables binaires	
	(1)	(2)	(3)
Log (PIB)	-1.449*** (0.005)	-1.920*** (0.000)	-1.502*** (0.005)
Log (pop 15-64 ans)	0.928 (0.408)	1.018 (0.399)	0.164 (0.884)
Mortalité infantile	-0.026 (0.363)	-0.021 (0.466)	-0.042 (0.170)
Log (TDbND)	0.063*** (0.001)	0.064*** (0.001)	0.063*** (0.001)
Intensité conflits	0.672*** (0.000)	-	-
Intensité (binaire) 1 si + de 25 morts	-	0.729*** (0.000)	-
Intensité (binaire) 1 si + de 999 morts	-	-	0.891*** (0.000)
$\alpha$ or	Oui	Oui	Oui
$\alpha$ rt	Oui	Oui	Oui
Pseudo R <sup>2</sup>	0.9387	0.9375	0.9376

Entre ( ) se trouvent les P>|z| valeurs. 17,148 observations. Source : construction par l'auteur.

Ici la spécification porte sur la variable qui mesure l'intensité des conflits puisqu'au lieu de lui donner 3 valeurs, on décide de la faire s'exprimer de manière binaire, de 2 façons différentes. Soit la valeur 1 signale que le conflit à causé plus de 25 morts, 0 si moins (1<sup>ère</sup> façon) ; soit la valeur 1 signale que le conflit à causé au moins 1000 morts (2<sup>ème</sup> façon).

On voit que les 2 façons donnent des coefficients tout aussi significatifs que la variable d'origine, mais que ceux-ci sont plus forts. C'est d'ailleurs la 2<sup>ème</sup> façon qui montre le plus grand effet, suggérant que c'est lorsque l'on passerait d'un conflit faible-moyen, à un conflit fort qu'on observerait le plus de détections. Les R<sup>2</sup> de ces 2 variables binaires sont un peu

moins élevés que celui de l'intensité d'origine, mais cela ne les rend pas inintéressants pour autant.

### 3.4. Régressions sur les effets fixes

A présent, les effets fixes  $\alpha_{ot}$  des régressions de Poissons ont été sauvegardés et on peut les utiliser comme nouvelle variable dépendante d'une régression classique avec les mêmes variables explicatives que pour les modèles sortis avec *ppmlhdfe*. Le but de cette manœuvre est de voir d'une autre manière si les variables choisies expliquent bien ce qui se trouve au cœur de l'effet fixe  $\alpha_{ot}$ . Ainsi, on peut tenter de voir si les 2 méthodes convergent et donnent des résultats similaires.

Tableau 8 : OLS sur les effets fixes récupérés ( $\alpha_{ot}$ )

	Benchmark	Sensibilité au retard			
	(1)	(2)	(3)	(4)	(5)
Log (PIB)	-0.282*** (0.000)	-0.296*** (0.000)	-0.297*** (0.000)	-0.302*** (0.000)	-0.298*** (0.000)
Log (pop 15-64 ans)	0.023** (0.034)	0.049*** (0.000)	0.062*** (0.000)	0.054*** (0.000)	0.032*** (0.003)
Mortalité infantile	-0.002*** (0.001)	-0.002*** (0.002)	-0.002*** (0.007)	-0.002*** (0.001)	-0.003*** (0.000)
Log (TDbND)	-0.034*** (0.000)	-0.079*** (0.000)	-0.102*** (0.000)	-0.072*** (0.000)	-0.038*** (0.000)
Intensité conflits	1.309*** (0.000)	1.321*** (0.000)	1.329*** (0.000)	1.300*** (0.000)	1.298*** (0.000)
R <sup>2</sup>	0.1398	0.1443	0.1477	0.1408	0.1362

Entre ( ) se trouvent les  $P > |z|$  valeurs. 26 950 observations. Source : construction par l'auteur.

Le tableau 8 doit donc être comparé au tableau 4 pour voir si les résultats des deux méthodes convergent.

A première vue, les résultats sont tous significatifs et du même signe que ceux de la méthode *ppmlhdfe*, sauf pour le log de TDbND qui montre un signe négatif avec la régression OLS. C'est un petit peu surprenant car jusqu'ici cette variable avait montré des effets positifs significatifs et résistants aux retards ou aux spécifications. Peut-être qu'il y a effectivement des effets multiples au sein de cette variable TDbND comme le suggère le tableau 6, où l'on observe un effet négatif et significatif des morts par cause de désastres biologiques alors que le nombre de victimes de désastres occasionnant de gros dommages montre l'opposé.

C'est la même idée qui créa le tableau 9 afin de le comparer au tableau 5. Après observation des deux, le raisonnement est moins évident puisque l'ajout des formes quadratiques fait changer le signe du log du PIB ainsi que de la mortalité infantile, ce qui n'était pas le cas avec la régression de Poisson. Ce n'est pas très grave car aucune confirmation de l'effet des formes

quadratiques n'était pas forcément attendue puisqu'aucune conclusion n'avaient été tirées du tableau 5 à ce propos. Cependant, dans le tableau 9, ici encore, la variable TDbND montre un effet constant, significatif et négatif, peu importe la nouvelle spécification. Heureusement, la variable qui mesure l'intensité des conflits est elle aussi constante, significative, et surtout du même signe que les résultats obtenus avec le premier modèle.

Tableau 9 : OLS avec spécification

	Benchmark	Variables <sup>2</sup>		Variables d'interaction	
	(1)	(2)	(3)	(4)	(5)
Log (PIB)	-0.282*** (0.000)	5.532*** (0.000)	-0.269*** (0.000)	-0.282*** (0.000)	-0.280*** (0.000)
[Log (PIB)] <sup>2</sup>	-	-0.338*** (0.000)	-	-	-
Log (pop 15-64 ans)	0.023** (0.034)	-0.027** (0.011)	0.019* (0.087)	0.025** (0.022)	0.029** (0.011)
Mortalité infantile	-0.002*** (0.001)	0.0001 (0.890)	0.002 (0.374)	-0.002*** (0.001)	-0.002*** (0.001)
[Mortalité infantile] <sup>2</sup>	-	-	-0.00005** (0.021)	-	-
Log (TDbND)	-0.034*** (0.000)	-0.029*** (0.000)	-0.033*** (0.000)	-0.030*** (0.000)	-0.032*** (0.000)
Intensité conflits	1.309*** (0.000)	1.408*** (0.000)	1.313*** (0.000)	1.307*** (0.000)	1.316*** (0.000)
TDbND*pop 15-64 ans	-	-	-	-1.48e-13 (0.069)	-
Intens.conflits*pop 15-64	-	-	-	-	-2.67e-10 (0.096)
R <sup>2</sup>	0.1398	0.1680	0.1399	0.1397	0.1399

Entre ( ) se trouvent les  $P > |z|$  valeurs. 26 950 observations. Source : construction par l'auteur.

Le dernier tableau 10, reprend le modèle « benchmark » de départ ainsi qu'un deuxième modèle qui rassemble les quelques variables alternatives qui in suscitent un intérêt. A savoir, les désastres détaillés en 3 types différents, et l'intensité des conflits mesurée de manière binaire (la 2<sup>ème</sup> façon).

A ces 2 modèles appelés « meilleures spécifications », on ajoute une dernière variable : les détections de la période  $t-1$ , donc le lag des détections observées. En effet, on veut tester si la possibilité qu'une même personne soit détectée plusieurs fois<sup>21</sup> a un impact sur nos résultats. On peut imaginer qu'un migrant illégal, arrêté et reconduit à l'extérieur des frontières ait l'envie de retenter sa chance et pourrait être détecté une seconde fois

<sup>21</sup> Frontex data as of 03 November 2017.

Et en effet, la variable Log (Détections t-1) est très significative et montre un effet positif non négligeable. Mais il peut aussi s'agir d'un effet « boule de neige », où la présence d'un migrant en temps  $t$  indique que le passage est « possible » ou en tout cas qu'il se tente. La variable  $t-1$  peut donc aussi avoir un effet sans forcément attendre que les mêmes individus passent plusieurs fois au même endroit. De plus, depuis l'intégration de cette dernière variable Log (Détections t-1), le Log (pop 15-64 ans) montre un effet positif et significatif à 95%, lui qui ne l'avait été que rarement jusqu'à présent.

### 3.5. Meilleures spécifications

Tableau 10 : meilleures spécifications

	Meilleures spécifications		Effet des détections t-1	
	(1)	(2)	(3)	(4)
Log (Détections t-1)	-	-	0.516*** (0.000)	0.518*** (0.000)
Log (PIB)	-1.449*** (0.005)	-1.710*** (0.001)	-0.400 (0.230)	-0.752** (0.045)
Log (pop 15-64 ans)	0.928 (0.408)	0.503 (0.641)	2.002** (0.018)	1.867** (0.027)
Mortalité infantile	-0.026 (0.363)	-0.046 (0.124)	0.027 (0.105)	0.019 (0.272)
Log (TDbND)	0.063*** (0.001)	-	0.073*** (0.000)	-
Log (TDbMétéo)	-	0.013 (0.638)	-	0.011 (0.676)
Log (TDbDestructif)	-	0.109*** (0.000)	-	0.108*** (0.000)
Log (TDbBio)	-	-0.070*** (0.005)	-	-0.036** (0.032)
Intensité conflits	0.672*** (0.000)	-	0.405*** (0.001)	-
Intensité (binaire) 1 si + de 999 morts	-	0.916*** (0.000)	-	0.261** (0.031)
$\alpha$ or	Oui	Oui	Oui	Oui
$\alpha$ rt	Oui	Oui	Oui	Oui
Pseudo R <sup>2</sup>	0.9387	0.9392	0.9622	0.9624

Entre ( ) se trouvent les  $P > |z|$  valeurs. 16,573 observations à cause de détections t-1.

Source : construction par l'auteur.

## 6. Conclusion

L'objectif de ce travail était d'analyser plus en détail l'un des 3 effets fixes qui expliquent le nombre de détections d'immigrants clandestins passant les frontières vers l'Europe. En choisissant de se focaliser sur l'effet fixe  $\alpha_{ot}$  (origine-temps) il fallait réfléchir aux variables qui l'expliqueraient le mieux, en volume et en qualité. Il fallait essayer d'intégrer ces nouvelles idées de variables tout gardant une base de données fournie avec un certain degré de précision temporelle. Une fois ce défi relevé, il fallait avoir les bons outils statistiques afin de mesurer leur impact tout en se protégeant de l'effet important des 2 premiers effets fixes.

Grace à ces outils, on a pu constater que certaines variables explicatives étaient bien choisies puisqu'elles sont restées significatives, constantes et robustes aux retards ainsi qu'à différentes spécifications, tout en montrant les mêmes types résultats convergents avec les 2 méthodes : *ppmlhdfe* et OLS sur les effets fixes. Il s'agit principalement de l'effet positif sur le nombre de détections induit par l'intensité des conflits armés dans le pays d'origine. Le PIB, pris en logarithme n'est pas loin, lui non plus, de donner tous les bons signaux d'un effet négatif significatif et robuste sur le nombre de détections. Seul l'ajout de sa forme quadratique perturbe un peu ses résultats, mais lors de cette régression, aucun des 2 coefficients n'était significatif. En revanche, on ne peut malheureusement pas conclure à un effet certain du nombre de victimes dues aux désastres naturels. Son effet est incertain et dépend de la méthode car les effets de cette variable, pourtant prometteuse, ne sont pas robustes à la méthode puisque les résultats ne convergent pas.

Ce travail atteint donc déjà ses limites puisqu'il n'arrive pas à expliquer pourquoi les effets cette variables, pourtant significatifs et robustes dans le premier modèle, divergent avec ses effets observés dans le second modèle.

Les perspectives de réponses pourraient être de creuser la nature de ces différents désastres naturels et d'essayer d'en extraire toutes les composantes, peut-être antinomiques.

Et afin d'étoffer encore ce travail, il aurait peut-être été intéressant d'essayer de trouver une ou deux autres variables explicatives recouvrant d'autres facettes de l'effet global origine-temps, afin d'en extraire la quasi-totalité de sa substance. Pourquoi pas une mesure de la qualité des institutions ou bien une mesure de la stabilité politique et du climat de sécurité-insécurité qui en dépend ? La difficulté réside aussi dans l'obligation pour cette nouvelle variable de correspondre à toutes les autres déjà présentes dans le modèle.

Ou alors, il est aussi possible de s'attaquer à l'un des deux autres effets fixes restants, puisqu'ils semblaient avoir même plus de poids que celui étudié ici. Cela nécessiterait d'autres ressources et d'autres esprits curieux.

## 7. Bibliographie

### Livres

MENDY T. (2009), L'immigration clandestine Mythes, mystères et réalités, l'Harmattan, Paris.

DUEZ D. (2008), L'Union européenne et l'immigration clandestine : de la sécurité intérieure à la construction de la communauté politique, Éditions de l'Université de Bruxelles.

### Bases de données

EM-DAT: The Emergency Events Database - Université catholique de Louvain (UCLouvain) - CRED, D. Guha-Sapir - [www.emdat.be](http://www.emdat.be), Brussels, Belgium;

Frontex data as of 03 November 2017.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand (2002) Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research* 39(5).

Högbladh Stina, 2019, "UCDP GED Codebook version 19.1", Department of Peace and Conflict Research, Uppsala University

International Labour Organization, ILOSTAT database. Data retrieved in November 2017.

Pettersson, Therese; Stina Högbladh & Magnus Öberg, 2019. Organized violence, 1989-2018 and peace agreements, *Journal of Peace Research* 56(4).

Sundberg, Ralph, and Erik Melander, 2013, "Introducing the UCDP Georeferenced Event Dataset", *Journal of Peace Research*, vol.50, no.4, 523-532

United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Population Prospects 2019*, custom data acquired via website.

(1) United Nations Population Division. *World Population Prospects: 2017 Revision*. (2) Census reports and other statistical publications from national statistical offices, (3) Eurostat: Demographic Statistics, (4) United Nations Statistical Division. *Population and Vital Statistics Reprot* (various years), (5) U.S. Census Bureau: *International Database*, and (6) Secretariat of the Pacific Community: *Statistics and Demography Programme*.

UN Inter-agency Group for Child Mortality Estimation (UNICEF, WHO, World Bank, UN DESA Population Division)

World Bank, *International Comparison Program database*.

World Bank national accounts data, and OECD National Accounts data files.

World Bank's total population and age/sex distributions of the United Nations Population Division's *World Population Prospects: 2017 Revision*.

World Development Indicators

## Articles

BEINE M., DOCQUIER F et ÖZDEN C. (2009), “Diasporas”, *Journal of Development Economics*.

BEINE M., DOCQUIER F et OZDEN C. (2015), “Dissecting network externalities in international migration”, *Journal of Demographic Economics*, n°81, pp379–408.

CORREIA S., GUIMARAES P et ZYLKIN TOM. (2019), “PPMLHDFE: Fast Poisson Estimation with High-Dimensional Fixed Effects”.

FRIEBEL G., MANCHIN M., MENDOLA M. et PRAROLO G. (2018), “Human Smuggling and Intentions to Migrate: Global Evidence from a Supply Shock along Africa-to-Europe Migration Routes”.

HARARI M et FERRERA E. (2018), “Conflict, Climate and Cells: A disaggregated analysis”, *Review of Economics and Statistics*, n°4(594-608)

HMJAK I., LUKIC T., GAVRILOV M., MARKOVIC S., UNKASEVIC M et TOSIC I. (2013), “Aridity in Vojvodina, Serbia”, *Theoretical and Applied Climatology*, vol 11, n°1-2, pp323-332.

## Pages internet

“Climate Change Knowledge Portal” (sans date), sur le site *World data bank group*. Consulté le 25 mai. 2019.

<https://climateknowledgeportal.worldbank.org/download-data>

“Congo Unemployment Rate” (sans date), sur le site *Trading economics*. Consulté le 15 juin. 2019.

<https://tradingeconomics.com/congo/unemployment-rate>

“Data catalog” (sans date), sur le site *world bank group*. Consulté le 13 juin.2019.

<https://datacatalog.worldbank.org/public-licenses#cc-by>

“DataBankWorld Development Indicators” (sans date), sur le site *World data bank group*. Consulté le 25 mai. 2019.

<https://databank.worldbank.org/indicator/NY.GDP.PCAP.CD/1ff4a498/Popular-Indicators#>

“Dominica - Mortality rate, infant (per 1,000 live births)” (sans date), sur le site *Trading economics*. Consulté le 15 juin. 2019.

<https://tradingeconomics.com/dominica/mortality-rate-infant-per-1-000-live-births-wb-data.html>

“Exploring 22 databases and 10 tables presentations” (sans date), sur le site *United Nations Statistics Division*. Consulté le 15juin.2019.

<http://data.un.org/Explorer.aspx>

“FAOSTAT” (sans date), sur le site *Food and Agriculture Organization of the United Nations*. Consulté le 27 mai.2019.

<http://www.fao.org/faostat/en/#data>

“Infant mortality rate worries health officials” (sans date), sur le site *vibes media group*. Consulté. le 26 mai.2019.

<https://www.dominicavibes.dm/news-80997/>

“Mortality rate, infant (per 1,000 live births)” (sans date), sur le site *Index Mundi*. Consulté le 10 juin.2019.

<https://www.indexmundi.com/facts/indicators/SP.DYN.IMRT.IN/compare#country=dm>

“UCDP Dataset Download Center” (sans date), sur le site *PRIO*. Consulté le 01 juin. 2019.

<https://ucdp.uu.se/downloads/#d3>

“UCDP/PRIO Armed Conflict Dataset” (sans date), sur le site *PRIO*. Consulté le 30 mai. 2019.

<https://www.prio.org/Data/Armed-Conflict/UCDP-PRIO/>

“UNICEF Data: Monitoring the situation of children and women” (sans date), sur le site *UNICEF*. Consulté le 10 mai.2019.

<https://data.unicef.org/country/dma/>

“Venezuela GDP per Capita”, (sans date), sur le site *CEIC*. Consulté le 25 mai.2019.

<https://www.ceicdata.com/en/indicator/venezuela/gdp-per-capita>

“Welcome to the EM-DAT website” (sans date), sur le site *EM-DAT The International Disaster Database*. Consulté le 16 juin.2019.

[www.emdat.be](http://www.emdat.be)

“World Population Prospects 2019” (sans date), sur le site *UNITED NATIONS DESA / POPULATION DIVISION*. Consulté le 02 mai. 2019

<https://population.un.org/wpp/DataQuery/>