

Note d'analyse n° 87  
« Création d'entreprise : de la volonté au passage à l'acte »

## ANNEXE TECHNIQUE

### 1. Présentation des trois modèles : avoir la volonté d'entreprendre, passer à l'acte et être un nouvel entrepreneur

La note d'analyse<sup>1</sup> porte sur l'impact des caractéristiques individuelles et des effets fixes pays sur la probabilité de devenir « un nouvel entrepreneur », l'objectif étant de décomposer cette probabilité entre « avoir la volonté d'entreprendre » et « passer à l'acte ».

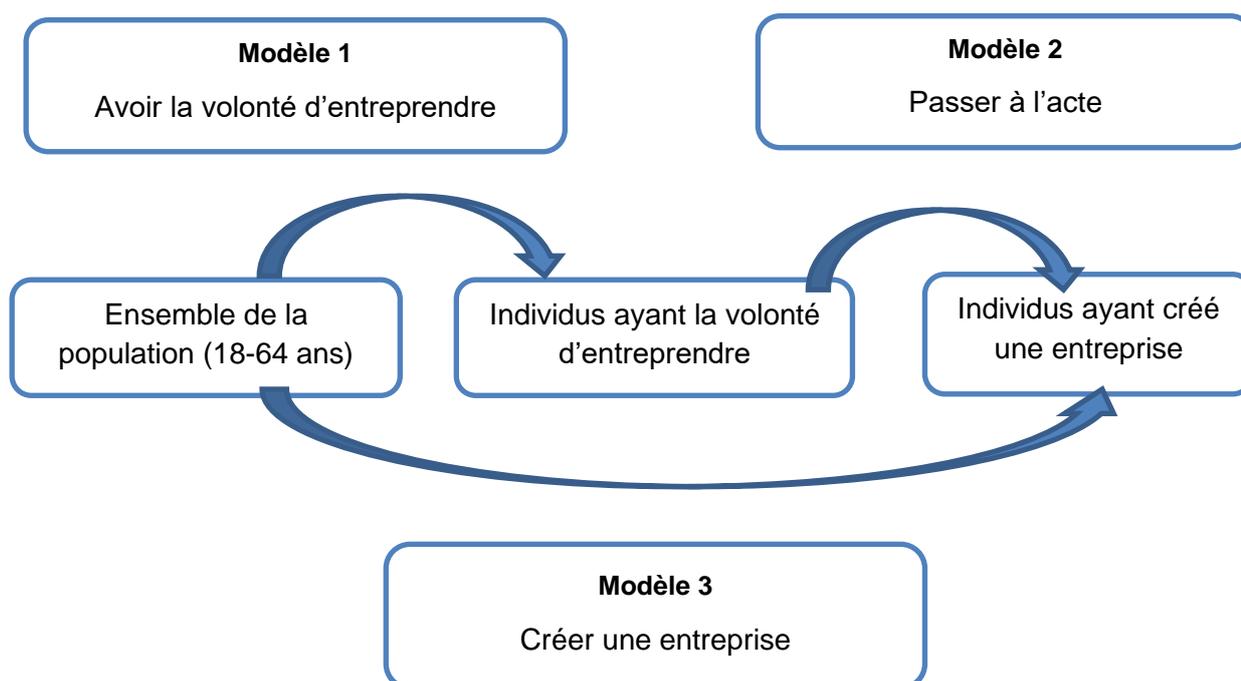
Le schéma 1 présente les trois étapes du processus entrepreneurial qui correspondent à trois états que l'on probabilise à l'aide d'un modèle logistique :

- le modèle logistique « **Avoir la volonté d'entreprendre** » évalue l'impact des caractéristiques individuelles, détaillées ci-après, sur la probabilité d'avoir envie de créer une entreprise dans les trois prochaines années (échantillon : ensemble de la population âgée de 18 à 64 ans) ;
- le modèle logistique « **Passer à l'acte** » évalue l'impact des caractéristiques individuelles sur la probabilité de réussir à créer une entreprise (être un « nouvel entrepreneur ») l'année n sachant que l'individu avait la volonté d'entreprendre l'année t-2 (échantillon : individus ayant la volonté d'entreprendre en période t-2). En l'absence de variable sur le « passage à l'acte » en t en ayant la volonté en t-2 dans le questionnaire source de la base GEM, ce modèle a conduit à créer un pseudo-panel sur une période de trois ans ;
- le modèle logistique « **Être un nouvel entrepreneur** » évalue l'impact des caractéristiques individuelles sur la probabilité d'avoir créé et de gérer une nouvelle entreprise depuis moins de trois ans et demi (échantillon : ensemble de la population âgée de 18 à 64 ans).

---

<sup>1</sup> Disponible sur le [site de France Stratégie](https://www.strategie.gouv.fr).

## Schéma 1 – Les étapes de l'entrepreneuriat dans l'étude



Source : France Stratégie

**Modèle économétrique :** le modèle utilisé est un modèle logistique multivarié car les variables dépendantes sont binaires (l'individu est un nouvel entrepreneur ou il ne l'est pas) et les variables indépendantes sont soit multiples, continues ou factorielles. Le choix du logit plutôt que du probit tient au fait qu'il est majoritairement utilisé dans la littérature pour l'estimation de la probabilité de devenir entrepreneur (Blanchflower et Oswald 1998<sup>2</sup>, Van der Zwan *et al.* 2010<sup>3</sup>), les deux méthodes présentant des résultats similaires sur nos données.

**Périodes étudiées :** l'ensemble des estimations avec les différentes spécifications a été réalisé sur les deux périodes étudiées, soit [2002, 2003, 2004] et [2012, 2013, 2014], ce qui permet une comparaison temporelle en plus d'une comparaison internationale, ainsi que de tenir compte des chocs institutionnels et conjoncturels ayant eu lieu entre-temps (création du statut de micro-entrepreneur, réformes de l'ACCRE<sup>4</sup> et crise financière de 2008). Le fait d'avoir choisi des périodes de trois années permet d'avoir une meilleure stabilité et robustesse des résultats en augmentant la taille de l'échantillon de 90 000 à 260 000 individus. Les coefficients des modèles logistiques sont plutôt stables entre 2002-2004 et 2012-2014, à l'exception de la France dont le rapport des cotes a été multiplié par 7 pour le modèle « avoir la volonté d'entreprendre » et par 2,5 pour le modèle « être un nouvel entrepreneur ». Le poids de l'auto-évaluation des compétences a également fortement

<sup>2</sup> Blanchflower D. G. et Oswald A. J. (1998), « What makes an entrepreneur? », *Journal of Labor Economics*, vol. 16 (1), janvier, p. 26-60.

<sup>3</sup> Van der Zwan P., Thurik R. et Grilo I. (2010), « The entrepreneurial ladder and its determinants », *Applied Economics*, vol. 42 (17), p. 2183-2191.

<sup>4</sup> ACCRE : aide à la création ou à la reprise d'entreprise.

augmenté entre les deux périodes pour l'ensemble des modèles, tout comme le fait de connaître des entrepreneurs dans son entourage proche.

**Spécification** : pour chacune des familles de modèles, trois spécifications différentes sont testées pour les régressions, une avec uniquement les variables socio-démographiques, une deuxième en ajoutant les variables psychologiques et sociales et une troisième en ajoutant les variables de perception de l'entrepreneuriat.

**Tableau 1 – Tableau de corrélation entre les variables du modèle**

	V.	N.	P.	G.	A.	E.	O.	C.	P.	E.	C.	V.	H.
Volonté	1												
Nouveau	0	1											
Pays (France)	0,01 (**)	-0,02 (***)	1										
Sexe	0,07 (***)	0,04 (***)	-0,01 (***)	1									
Age	-0,15 (***)	-0,03 (***)	0,01 (***)	-0,03 (***)	1								
Éducation	0,04 (***)	0,04 (***)	-0,03 (***)	-0,02 ( )	-0,08 (***)	1							
Opportunité	0,17 (***)	0,09 (***)	-0,06 (***)	0,06 (***)	-0,06 (***)	0,11 (***)	1						
Compétences	0,21 (***)	0,16 (***)	-0,04 (***)	0,15 (***)	0,03 (***)	0,11 (***)	0,13 (***)	1					
Peur	-0,09 (***)	-0,08 (***)	-0,01 ( )	-0,09 (***)	-0,02 (***)	-0,02 (***)	-0,13 (***)	-0,14 (***)	1				
Entourage	0,16 (***)	0,16 (***)	-0,02 ( )	0,07 (***)	-0,1 (***)	0,1 (***)	0,18 (***)	0,23 (***)	-0,05 (***)	1			
Carrière désirable	0,06 (***)	0,01 (*)	0,02 (***)	0,01 (*)	-0,03 (***)	-0,1 (***)	0,07 (***)	0,00 ( )	-0,01 (**)	-0,01 (**)	1		
Vision positive media	0,04 (***)	0,02 (***)	-0,04 (***)	0,01 (*)	0,05 (***)	-0,00 (.)	0,12 (***)	0,019 (***)	-0,02 (***)	0,02 (***)	0,13 (***)	1	
Haut statut social	0,01 (***)	-0,01 (***)	0,03 (***)	-0,00 ( )	-0,02 (***)	0 ( )	0,08 (***)	-0,04 (***)	0,04 (***)	-0,02 (***)	0,16 (***)	0,14 (***)	1

Période : 2012-2014.

Échantillon : OCDE, individus de 18 à 64 ans.

Source : GEM, calculs France Stratégie

### La robustesse des résultats a été contrôlée par plusieurs tests différents :

- Utilisation d'une régression probit ou MCO plutôt que logit.
- Estimation des modèles en enlevant les variables une à une et en regardant comment évoluent les coefficients des régressions ainsi que le R2 ajusté. Globalement le signe des coefficients estimés ne change jamais et les résultats obtenus sont proches des modèles retenus.
- Analyse de la robustesse par *bootstrap* : découpage de l'échantillon en un ensemble de sous-échantillons et estimation des coefficients pour l'ensemble des sous-échantillons.
- Comparaison des résultats obtenus après plusieurs itérations.
- La manière dont la variable « éducation » a été renseignée dans GEM étant différente en fonction du système éducatif des pays et de la période, la robustesse de la variable « éducation » a été testée en regroupant les 5 modalités en 3 modalités (« moins que le baccalauréat », « baccalauréat » et « plus que le baccalauréat »). Les résultats sont robustes à l'agrégation de modalités, on obtient des coefficients similaires et la significativité de la variable n'est pas impactée.

### Encadré 1 – Exemple pour la famille de modèle « avoir la volonté d'entreprendre »

Notre échantillon est composé de 264 139 individus âgés de 18 à 64 ans, dont certains ont la volonté d'entreprendre et pas d'autres. On cherche à estimer la probabilité qu'un individu ait la volonté d'entreprendre, soit la probabilité que la variable dépendante  $Y$  soit égale à 1 sachant  $X_i$ , soit  $P(Y_i = 1 | X_i)$ .

$$Y_i = \begin{cases} 1 & \text{si l'individu a la volonté d'entreprendre} \\ 0 & \text{si l'individu n'a pas la volonté d'entreprendre} \end{cases} = \mathbb{1}_{X_i\theta + \varepsilon_i > 0}$$

Chaque individu présente  $K$  caractéristiques liées à son origine géographique (son pays), ses caractéristiques socio-démographiques (âge, sexe, éducation), psychologiques et sociales (perception d'opportunités entrepreneuriales, auto-perception des compétences, peur de l'échec, présence d'un entourage entrepreneurial), ainsi que trois variables liées à la perception de l'entrepreneuriat. Chaque modèle est réalisé avec trois spécifications différentes, ainsi  $K$  varie en fonction du modèle. Les variables ont été sélectionnées suite à la réalisation d'une revue de littérature<sup>5</sup>.

La probabilité qu'un individu ait la volonté d'entreprendre est estimée selon la formule indiquée ci-dessous, avec  $K$  le nombre de variables indépendantes du modèle,  $\beta_i$  les coefficients obtenus pour chacune des  $K$  variables explicatives,  $\beta_0$  la constante et  $u_i$  le terme d'erreur.

$$Y_i = \beta_0 + \sum_{i=0}^K \beta_i X_i + u_i$$

Le modèle Logit correspond à la loi logistique de fonction de répartition  $\Lambda$  :

$$F(X_i\beta) = \Lambda(X_i\beta) = \frac{e^{X_i\beta}}{1 + e^{X_i\beta}} = \frac{1}{1 + e^{-X_i\beta}}$$

Une fois le modèle estimé, on obtient des valeurs pour les paramètres  $\beta_i$  dont on ne peut interpréter que le signe. L'aspect essentiel est l'effet marginal de la  $i$ -ème variable  $X_{ij}$  sur la probabilité de l'évènement  $Y=1$  pour l'individu  $i$ . Cet effet s'écrit pour une variable  $X_{ij}$  continue (pour une variable explicative qualitative, il faut considérer un taux d'accroissement) :

$$\frac{\partial F(X_i\beta)}{\partial X_{ij}} = f(X_i\beta)\beta_j$$

Si  $\beta_j > 0$ ,  $X_{ij}$  a un effet positif sur la probabilité que l'individu ait la volonté d'entreprendre. Sinon, cet effet est négatif.

Pour évaluer l'ampleur de l'effet d'une variable, on s'appuie sur l'interprétation des *odds ratios* ou rapport des cotes. Les *odds ratios* s'interprètent comme la variation de la probabilité d'avoir la volonté d'entreprendre liée à la variable étudiée, toutes choses égales par ailleurs. Par exemple, être une femme réduit de 18 % la probabilité d'avoir la volonté d'entreprendre.

<sup>5</sup> Bergmann H., Mueller S. et Schrettle T. (2014), « The use of global entrepreneurship monitor data in academic research: A critical inventory and future potentials », *International Journal of Entrepreneurial Venturing*, vol. 6 (3), janvier, p. 242-276.

$$OR_1 = \frac{\beta_1 / (1 - \beta_1)}{\beta_0 / (1 - \beta_0)} \text{ avec } \beta_0 \text{ la modalité de référence pour une variable catégorielle.}$$

Plusieurs modèles peuvent être comparés sur les mêmes données, le meilleur modèle est celui pour lequel les critères d'Akaike ou de Schwarz sont les plus petits, et le R2 ajusté est le plus élevé.

Dans la dernière partie de la note d'analyse, nous réalisons des calculs de contributions de chaque variable à la probabilité d'entreprendre dans quatre pays (France, Allemagne, États-Unis, Royaume-Uni) et nous expliquons les différences entre la France et les trois autres pays sélectionnés. Cette décomposition a été réalisée en multipliant les coefficients du modèle par les valeurs moyennes des variables indépendantes pour chacun des quatre pays, puis obtention des résidus par différence avec la valeur observée. Les différences ont ensuite été calculées pour

chaque pays :  $\beta_i X_i^{FR} - \beta_i X_i^{USA}$ , de même avec les résidus. Enfin, le passage aux probabilités pour une interprétation aisée des résultats a été réalisé avec la formule du logit :

$$p = \frac{e^{\alpha + \beta_0 X_0 + \beta_1 X_1 + \dots + \beta_n X_n}}{1 + e^{\alpha + \beta_0 X_0 + \beta_1 X_1 + \dots + \beta_n X_n}}$$

**Tableau 2 – Spécifications testées pour les modèles économétriques**

Famille de variables	Variables	Type de variables	Spécifications			
			n° 1	n° 2	n° 3	n° 4
Effet fixe pays	Pays	Catégorielle (37)	x	x	x	x
	Genre	Binaire	x	x	x	x
Variables socio-démographiques	Âge	Catégorielle (5)	x	x	x	x
	Éducation	Catégorielle (5)	x	x	x	x
	Statut professionnel	Catégorielle (6)				x
Variables psychologiques et sociales	Opportunité	Binaire		x	x	x
	Compétences	Binaire		x	x	x
	Peur	Binaire		x	x	x
	Entourage	Binaire		x	x	x
Variables de perception de l'entrepreneuriat	Carrière désirable	Binaire			x	x
	Vision positive media	Binaire			x	x
	Haut statut social	Binaire			x	x

Source : France Stratégie

En se basant sur les résultats obtenus, la spécification retenue pour les modèles « volonté d'entreprendre » et « être un nouvel entrepreneur par opportunité » est la spécification n° 3. Pour les modèles « passage à l'acte » et « être un nouvel entrepreneur », la spécification retenue est la numéro 2.

**Tableau 3 – Résultats économétriques des trois modèles  
pour la spécification n° 3 (toutes les variables)**

		Volonté d'entreprendre	Passage à l'acte	Nouvel entrepreneur	Entrepreneur par opportunité
Pays	France	1,52**	0,20***	0,53***	1,00
	États-Unis	Ref	Ref	Ref	Ref
	Royaume-Uni	0,56***	7,03***	1,13	0,56***
	Allemagne	0,82*	1,31	0,76***	0,77
Genre	Homme	Ref	Ref	Ref	Ref
	Femme	0,82***	1,02	0,81***	0,83
Âge	18-24 ans	1,83***	0,28***	0,68***	1,56***
	25-34 ans	1,31***	0,60***	1,04	1,09.
	35-44 ans	Ref	Ref	Ref	Ref
	45-54 ans	0,77***	1,03	0,81***	0,98
	55-64 ans	0,44***	0,55***	0,54***	0,95
Éducation	Aucun diplôme	Ref	Ref	Ref	Ref
	Collège	0,94	0,75.	0,88	1,65***
	Baccalauréat	0,99	0,81	0,92	2,01***
	Licence	1,08	0,99	0,99	2,55***
	Master	1,16**	0,97.	1,05	3,21***
Variables psychologiques et sociales	Opportunité	1,54***	1,04	1,44***	1,31***
	Compétences	2,94***	1,36***	4,95***	1,05
	Peur	0,77***	0,80***	0,60***	0,93.
	Entourage	1,61***	1,65***	3,64***	0,82***
Vision de l'entrepreneuriat	Carrière désirable	1,21***			0,94
	Vision positive media	1,09***			1,02
	Haut statut social	1,06***			1,12***
	R2	0,24	0,15	0,36	0,36

Période : 2012-2014.

Échantillon : OCDE, individus de 18 à 64 ans.

Source : GEM, calculs France Stratégie

À partir des modèles logistiques présentés précédemment, il est possible **d'estimer pour chaque étape du processus entrepreneurial si l'effet de composition de la population explique la différence de performance entre la France et les trois pays** de comparaison que sont les États-Unis, le Royaume-Uni et l'Allemagne (Partie III de la Note d'analyse). Pour cela, on simule pour chaque pays la probabilité que les individus constituant sa population soient à un certain stade de l'entrepreneuriat à partir des caractéristiques individuelles moyennes dans le pays, et sans tenir compte de leur pays d'origine<sup>6</sup>. En testant l'existence de différences entre les probabilités simulées pour chaque pays (quand on ne

<sup>6</sup> Appréhendé dans les modèles par un effet fixe pays. Ce dernier « capture » l'ensemble des facteurs propres à chaque pays (institutionnels, conjoncturels, etc.) autres que ceux décrits par les variables explicatives des modèles relatives aux caractéristiques socio-démographiques, psychologiques et de perception de l'entrepreneuriat. La significativité d'un effet de composition de la population expliquant une partie de la différence de probabilité d'être à différents stades de l'entrepreneuriat est appréciée à travers un test de Welch réalisé sur les prédictions d'être à un certain stade de l'entrepreneuriat quand on ne tient pas compte de l'effet pays lors de la prédiction. Si on observe une différence significative entre les prédictions de deux pays, alors on accepte que l'effet de composition explique au moins partiellement la différence entre les deux pays.

prend pas en compte les effets fixes pays), on identifie que l'effet de composition de la population explique significativement une part de ces différences pour la volonté d'entreprendre (modèle 1) et la probabilité d'être un nouvel entrepreneur (modèle 3). En revanche, l'effet de composition n'explique pas de manière significative la différence de probabilité de passer à l'acte entre la France et les autres pays de l'étude, à l'exception des États-Unis. Cette différence s'explique ainsi entièrement par d'autres facteurs (tableau 4).

**Tableau 4 – Test d'un effet de composition pour expliquer le fait d'être à différents stades de l'entrepreneuriat en France par rapport à d'autres pays**

	Volonté d'entreprendre	Passer à l'acte	Être un nouvel entrepreneur
États-Unis	0,001	0,001	0,001
Royaume-Uni	0,001	0,240	0,001
Allemagne	0,001	0,057	0,001
OCDE	0,001	0,811	0,001

Note de lecture : si la valeur (p-valeur) est < 0.05, alors il existe un effet de composition au seuil de tolérance de 5 %, Si au contraire la valeur (p-valeur) est > 0.05, alors on ne peut pas conclure à l'existence d'un effet de composition pour expliquer la différence de performance entre la France et le pays correspondant. Par exemple, pour la volonté d'entreprendre, la valeur est de 0,001 pour la comparaison avec les États-Unis. On conclut à l'existence d'un effet de composition.

Champ : individus âgés de 18 à 64 ans. 2012-2014.

Source : GEM, calculs France Stratégie

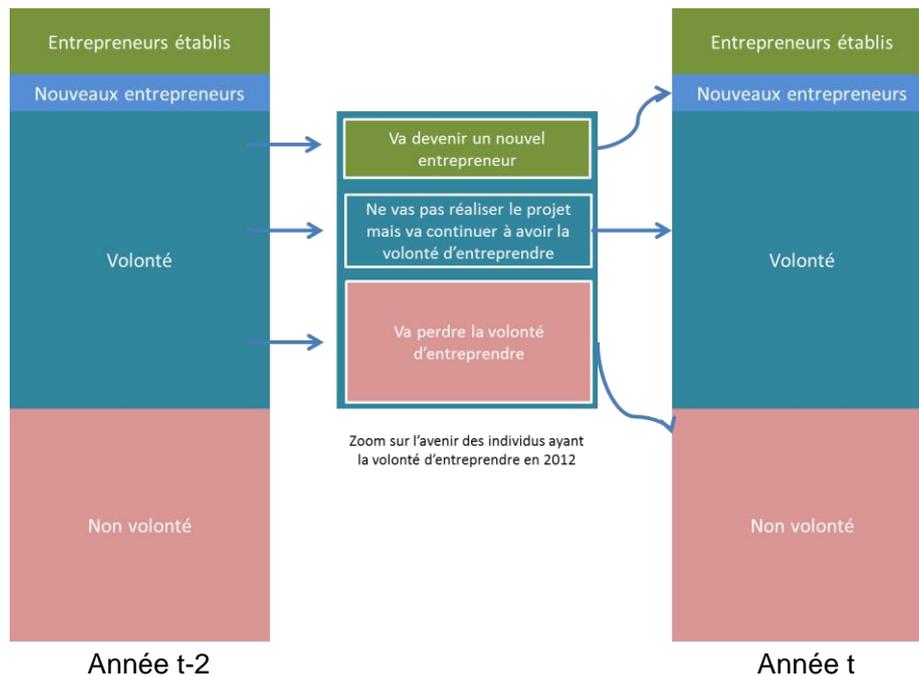
## 2. Présentation de la méthode économétrique pour le modèle « passer à l'acte »

Puisque l'enquête GEM est une enquête avec données en coupe et non en panel, on ne peut pas suivre le même individu sur plusieurs années. Or l'enjeu du modèle « passer à l'acte » est de réussir à retrouver parmi les individus ayant la volonté d'entreprendre en période t-2 les individus ayant réussi à créer leur entreprise en période t (deux ans après). Il faut donc décomposer la population des individus ayant envie d'entreprendre en période t-2, entre ceux qui arriveront à réaliser leur projet entrepreneurial et ceux qui ne le réaliseront pas. Cela fait, il est alors possible d'analyser l'impact des caractéristiques individuelles sur le passage à l'acte.

On cherche ainsi à estimer :

$$P(\text{'nouvel entrepreneur en } t' = 1 \mid \text{'volonté d'entreprendre en } t - 2' = 1)$$

## Graphique 1 – Que deviennent en t-2 les individus ayant la volonté d’entreprendre en t ?



Lecture : un individu ayant la volonté d’entreprendre en t-2 peut être à trois stades différents en t : il peut soit avoir perdu sa volonté d’entreprendre, soit avoir gardé sa volonté d’entreprendre mais ne pas avoir réalisé son projet entrepreneurial, soit avoir réussi à réaliser son projet entrepreneurial (et ainsi être devenu un « nouvel entrepreneur »).

Source : France Stratégie

Pour étudier la première transition des individus ayant la volonté d’entreprendre en n-2 à ceux qui y arriveront versus ceux qui échoueront/abandonneront en n, et pour pallier l’absence de données de panel, nous avons eu recours à la constitution d’un pseudo-panel sur deux périodes en s’inspirant de la méthode de Deaton (1985)<sup>7</sup>. Le principe des pseudo-panels est de suivre dans le temps non plus des individus, mais des cohortes, c’est-à-dire des groupes d’individus partageant un ensemble de caractéristiques les plus proches possibles pour toutes les périodes.

L’échantillon étudié en période t est déjà défini, il s’agit des « nouveaux entrepreneurs ». Pour retrouver l’échantillon en période t-2, de sorte que les deux groupes soient les plus semblables possibles, une des solutions est de retenir l’appariement sur score de propension comme le présentent Seawright (2009)<sup>8</sup> ou encore Chih-Lin *et al.* (2011)<sup>9</sup>. Cette méthode propose de trouver pour chaque « nouvel entrepreneur » en période t, son plus proche jumeau parmi les individus ayant la volonté d’entreprendre en t-2. L’appariement sur score de propension se fait en deux étapes : d’abord une estimation des scores de propension (généralement réalisée en utilisant un probit avec comme variable dépendante

<sup>7</sup> Deaton A. (1985), « Panel data from time series of cross-sections », *Journal of Econometrics*, vol. 30 (1-2), p. 109-126.

<sup>8</sup> Seawright J. (2009), « Matching for pseudo-panel inference », *Working Paper*, août

<sup>9</sup> Li Chih-Lin.C-L. (2011), « Propensity score matching in observational studies with multiple time points », Diss., The Ohio State University.

« être un nouvel entrepreneur ») et deuxièmement un appariement sur les scores estimés en fonction d'une procédure que l'on définit ci-dessous.

## 2.1. Estimation des scores de propension

La méthode d'estimation des scores de propension a été réalisée au moyen d'une forêt aléatoire, qui est un algorithme de *machine learning*. Les résultats obtenus avec le probit n'étaient pas satisfaisants (le modèle n'arrivait pas à correctement identifier qui étaient les nouveaux entrepreneurs avec un taux de bonnes réponses satisfaisant) et de surcroît cela aurait signifié que nous aurions utilisé le même modèle pour l'étape d'estimation des scores de propension et l'étape d'estimation de la probabilité de devenir entrepreneur quand on a la volonté en période n-2.

Le manuel de Faraway (2016)<sup>10</sup> indique qu'un modèle logistique n'est pas toujours le plus adapté pour estimer des scores de propension dans le cas où les relations entre variables dépendantes et indépendantes ne sont pas linéaires et difficilement modélisables avec un modèle logistique. Un modèle de forêt aléatoire peut être plus adapté. L'algorithme de forêt aléatoire effectue un apprentissage sur de multiples arbres de décision entraînés sur des sous-ensembles de données légèrement différents. L'utilisation d'arbre de décision ou de forêt aléatoire pour l'estimation de score de propension est de plus en plus répandue dans la littérature (Abt et McBride, 2017<sup>11</sup> ; Zhang, 2017<sup>12</sup> ; Cannas et Arpino, 2019<sup>13</sup>).

Pour l'estimation des scores de propension par forêt aléatoire, l'ensemble des variables disponibles ont été introduites, l'algorithme a sélectionné les plus pertinentes. Il en sort que les variables pays, entourage, compétences et âge sont les plus utiles pour prédire le fait de devenir un nouvel entrepreneur.

## 2.2. Réalisation de l'appariement sur score de propension

Une fois les scores de propension estimés, on peut identifier les individus qui se ressemblent le plus et ainsi réaliser l'appariement. La méthode est inspirée du papier de Zhao *et al.* (2016)<sup>14</sup> avec intégration des critiques de King et Nielsen (2016)<sup>15</sup> à l'encontre de l'utilisation de l'appariement sur score de propension. Cette critique porte principalement sur sa faiblesse à approximer un essai randomisé contrôlé et la déformation des caractéristiques de l'échantillon dans certains cas plutôt que l'amélioration de la représentativité.

---

<sup>10</sup> Faraway J. J. (2016), *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC.

<sup>11</sup> Nichols A. et McBride L. (2017), « Propensity scores and causal inference using machine learning methods », Stata Conference 13, Stata Users Group.

<sup>12</sup> Zhang Z. (2017), « Propensity score method: A non-parametric technique to reduce model dependence », *Annals of Translational Medicine*, vol. 5 (1).

<sup>13</sup> Cannas M. et Arpino B. (2019), « A comparison of machine learning algorithms and covariate balance measures for propensity score matching and weighting », *Biometrical Journal*, vol. 61 (2), mai.

<sup>14</sup> Zhao P., Su X., Ge T. et Fan J. (2016), « Propensity score and proximity matching using random forest », *Contemporary Clinical Trials*, vol. 47, mars, p. 85-92.

<sup>15</sup> King G. et Nielsen R. (2019), « Why propensity scores should not be used for matching », *Political Analysis*, vol. 27 (4), p. 1-20.

Les auteurs conseillent de limiter la distance maximale entre les scores de propension de deux individus appariés pour éviter une trop grande déformation des caractéristiques des individus des deux échantillons.

Les auteurs conseillent également l'identification de l'ordre optimal dans lequel les individus sont appariés. Après de multiples essais, la méthode offrant le meilleur résultat fut l'appariement aléatoire combiné à un algorithme de la distance entre l'échantillon d'origine et celui reconstitué. Dix mille appariements différents ont été générés et celui ayant les caractéristiques les plus similaires au groupe d'origine a été sélectionné pour le reste de l'analyse.

Les auteurs conseillent également de réaliser du *full matching* sur certaines variables clés de l'étude afin d'obtenir un appariement de meilleure qualité. Cela consiste à indiquer au logiciel que l'on souhaite que les individus appariés aient une ou plusieurs caractéristiques en commun quoi qu'il en soit. Nous avons appliqué du *full matching* sur les variables pays et âge, qui sont centraux dans notre étude.

Au terme de ces différentes étapes, on obtient un sous-échantillon parmi les individus ayant la volonté d'entreprendre en période t-2 particulièrement similaire au groupe des nouveaux entrepreneurs en période t. La distribution des scores de propension est quasiment identique entre les deux groupes, de même que la distribution des caractéristiques individuelles. Les individus ayant la volonté d'entreprendre en période t-2 faisant partie de ce sous-échantillon obtiennent alors la valeur 1 à la variable « passer à l'acte », alors que les individus ne faisant pas partie de ce sous-échantillon obtiennent la valeur 0 à cette variable. On réalise enfin une régression logistique sur notre nouvel échantillon et on estime la probabilité de passer à l'acte, c'est-à-dire la probabilité de devenir un nouvel entrepreneur quand on avait la volonté d'entreprendre deux ans auparavant.

**Tableau 4 – Statistiques descriptives OCDE**

		2002-2004						2012-2014					
		Ensemble	18-24	25-34	35-44	45-54	55-64	Ensemble	18-24	25-34	35-44	45-54	55-64
Genre	Homme	50,4	51,15	50,83	50,74	50,05	49,21	50,01	50,94	50,46	50,33	49,79	48,65
	Femme	49,6	48,85	49,17	49,26	49,95	50,79	49,99	49,06	49,54	49,67	50,21	51,35
Éducation	Aucun diplôme	0,16	0,03	0,03	0,12	0,23	0,36	4,18	2,51	2,19	2,92	4,78	8,71
	Collège	28,87	25,25	1,88	26,71	31,54	40,28	19,51	18,42	15,42	18,18	21,75	24,27
	Baccalauréat	36,72	45,10	35,01	37,10	35,72	32,82	37,75	55,87	34,43	35,50	36,15	33,11
	Licence	23,19	23,99	28,82	24,03	21,26	12,65	30,99	20,51	37,59	34,70	29,95	27,30
	Master	11,06	5,63	14,19	12,04	11,26	9,90	7,57	2,69	10,37	8,70	7,37	6,61
Variables psychologiques et sociales	Opportunité	34,16	34,76	37,33	34,92	32,43	30,65	31,95	36,43	34,71	31,41	29,72	28,51
	Compétences	45,9	33,35	38,19	50,95	48,04	43,37	44,96	32,06	47,24	49,87	48,26	41,58
	Peur	34,78	33,92	36,72	36,45	35,02	30,29	46,51	43,81	48,46	48,65	46,56	43,31
	Entourage	37,17	40,25	43,62	39,56	33,36	26,61	32,85	33,02	40,3	35,6	29,79	23,89
Vision de l'entrepreneuriat	Carrière désirable	59,07	67,08	60,96	57,12	56,11	56,4	56,11	61,45	56,2	53,91	54,2	57,12
	Vision positive media	51,85	49,5	51,12	51,53	52,7	54,04	51,14	48,66	49,31	50,52	52,44	54,54
	Haut statut social	64,54	71,1	65,61	62,65	61,74	63,05	64,71	69,57	65,01	62,59	63,21	65,23
Devenir entrepreneur	Volonté	9,78	14,17	13,54	10,26	7,51	3,86	13,94	20,93	18,49	14,37	11,1	6,32
	Émergent	3,4	3,27	4,45	4,06	3,12	1,67	4,86	4,57	6,35	5,53	4,63	2,67
	Nouveau	2,81	1,82	3,9	3,49	2,66	1,48	3,36	2,37	4,5	4,18	3,22	1,84
	Établi	5,64	0,79	3,4	7,04	8,17	7,36	7,14	1,14	3,74	8,36	10,78	9,74
Part nouveau opportunité		81 %	82,8 %	81,4 %	81 %	79 %	75,1 %	80 %	85 %	81 %	79 %	75 %	77 %
Rapport nouveau / volonté		0,29	0,13	0,29	0,34	0,35	0,38	0,24	0,11	0,24	0,29	0,29	0,29

Échantillon : 29 pays de l'OCDE, individus âgés de 18 à 64 ans.

Note de lecture : en 2002-2004 dans l'OCDE, 14,17 % des 18-24 ans avaient la volonté d'entreprendre dans les trois prochaines années.

Source : GEM, calculs France Stratégie