

Comprendre le changement social et les différences entre pays avec les modèles multiniveaux. L'exemple des attitudes économiques dans la *European Values Study*.

Frédéric Gonthier, Sciences Po Grenoble-PACTE, Université de Grenoble

Si les travaux statistiques analysant les relations entre les individus et leur environnement remontent aux années 60, l'apparition de méthodes dédiées à la modélisation des effets de contexte est plus récente. Il faut attendre le milieu des années 80 pour que se développent les modèles dits multiniveaux, ou hiérarchiques linéaires. Modèles mixtes conçus pour analyser des données emboîtées les unes dans les autres et présentant une structure hiérarchisée, ils ont surtout été utilisés en sciences sociales dans le domaine des sciences de l'éducation. Leur application à la sociologie et à la science politique reste aujourd'hui plus limitée.

L'objectif de cette présentation est de mettre en évidence les apports et les limites des modèles multiniveaux à l'analyse des données d'opinion issues des grandes enquêtes sociales internationales. Précisément, en mobilisant les trois dernières vagues de l'enquête sur les valeurs des Européens, on montre comment ces modèles permettent de mieux saisir l'évolution des attitudes économiques des Européens, en particulier l'évolution de la demande d'Etat parmi les pays membres de l'Union européenne.

Les données analysées

Le programme d'enquête sur les valeurs des Européens (*European Values Study*) interroge les attitudes, les opinions, les préférences et les croyances des individus à travers toute l'Europe. Plusieurs thèmes sont explorés : famille, travail, religion, politique, économie, immigration, environnement, relations sociales, engagements associatifs... Quatre vagues ont été réalisées entre 1981 et 2008 dans un nombre croissant de pays avec un questionnaire standardisé. 47 pays ont été enquêtés en 2008.

L'analyse présentée porte sur un jeu de données incluant les 23 pays de l'Union européenne présents dans les trois dernières vagues d'enquête (1990, 1999 et 2008) : Allemagne (Ouest+Est), Autriche, Belgique, Bulgarie, Danemark, Espagne, Estonie, Finlande, France, Hongrie, Irlande, Italie, Lettonie, Lituanie, Pays-Bas, Pologne, Portugal, République Tchèque, Roumanie, Royaume-Uni (Grande Bretagne + Irlande du Nord), Slovaquie, Slovénie, Suède.

Le jeu de données utilisé est composé d'échantillons nationaux de taille égale (environ 1000 individus par pays pour chaque vague), et est exploité sans pondération afin de ne pas biaiser l'analyse en donnant davantage de poids aux pays comportant le plus d'habitants (comme l'Allemagne, la France ou le Royaume-Uni). Le jeu de données contient un total de 68472 individus.

Quelques avantages des modèles multiniveaux pour l'analyse des enquêtes internationales

Les modèles multiniveaux permettent une meilleure estimation des différences entre pays. Alors que les analyses de régression "classiques" (méthode des moindres carrés ordinaire) reposent sur des hypothèses contraignantes quant à l'indépendance des termes d'erreur et l'homoscédasticité, l'analyse multiniveaux tient à la fois compte de la non-indépendance des observations et de l'hétérogénéité entre les termes d'erreur.

Dans le modèle MCO, les résidus sont considérés comme indépendants les uns des autres. Cette hypothèse d'indépendance des erreurs n'est plus forcément vérifiée lorsqu'on suppose l'existence d'effets de groupe. Autrement dit, quand les individus d'un même groupe sont proches sur un ensemble de variables (dont certaines vont demeurer inobservées), l'hypothèse d'indépendance des erreurs n'est pas validée. Cela affecte la précision dans les estimations du modèle, en particulier les erreurs-types des coefficients estimés. L'analyse multiniveaux permet une spécification plus complexe des résidus, qui sont en fait estimés à deux niveaux (en l'occurrence, au niveau 1 des individus et au niveau 2 des pays).

D'autre part, le modèle des MCO suppose l'homoscédasticité des erreurs (les erreurs auraient une variance constante). Les modèles multiniveaux remplacent cette hypothèse par un postulat plus souple : la variance des résidus peut varier comme une fonction linéaire ou non des variables explicatives. Autrement dit, il peut exister de l'hétérogénéité dans les relations entre les

groupes. Au lieu de contraindre la variance des résidus à être constante, l'analyse multiniveaux permet donc de les faire varier selon différentes variables explicatives. On peut ainsi mesurer l'hétérogénéité des comportements selon le contexte : en l'occurrence, les comportements moyens varient-ils selon les pays ? Les effets des variables explicatives varient-ils selon les pays ?

Construction de la variable dépendante « demande d'Etat »

Les enquêtes Valeurs interrogent les préférences économiques des Européens sous la forme d'échelles en 10 points opposant deux orientations contraires, l'une typique d'une position libérale, l'autre d'une position plutôt interventionniste. Elles sont formulées comme suit :

- « Les individus devraient avoir davantage la responsabilité de subvenir à leurs propres besoins » vs « L'Etat devrait avoir davantage la responsabilité d'assurer à chacun ses besoins ».
- « La propriété privée des entreprises et des industries devrait être développée » vs « La nationalisation des entreprises et des industries devrait être développée »
- « La concurrence est une bonne chose. Elle pousse les gens à travailler dur et à trouver de nouvelles idées » vs « La concurrence est dangereuse. Elle conduit à développer ce qu'il y a de pire chez les gens »
- « Les chômeurs devraient être obligés d'accepter tout emploi disponible ou bien perdre leur indemnité de chômage » vs « Les chômeurs devraient avoir le droit de refuser un emploi qui ne leur convient pas »
- « Il faudrait encourager davantage les efforts individuels » vs « Les revenus devraient être plus égalitaires »

Pour apprécier dans quelle mesure ces préférences s'organisent autour de dimensions distinctes, une analyse en composantes principales a été réalisée sur le fichier cumulant les données EVS de 1990, 1999 et 2008, à l'échelle de 23 pays de l'Union européenne. L'ACP permet ici de repérer la structure sous-jacente aux attitudes économiques des Européens. Deux dimensions se dégagent après une rotation de type Varimax (**figure 1**). La première recouvre les questions concernant le contrôle de l'Etat sur l'économie. Cette dimension est la plus saillante puisqu'elle rend compte de 30 % de la variabilité du nuage de points. Les items liés à l'égalité renvoient à une seconde dimension, qui explique 19 % de la variance et qui est surtout structurée par l'opposition entre les répondants qui considèrent que les revenus devraient être plus égaux et ceux qui pensent que les efforts individuels devraient être mieux récompensés¹.

La variable dépendante utilisée dans la suite de l'analyse correspond aux coordonnées/corrélations factorielles des individus sur le premier facteur de l'ACP. Elle indique donc la plus ou moins grande demande d'Etat de la part des Européens. Son amplitude a été ramenée à des variations comprises en 0 et 1, afin de pouvoir lire plus facilement les coefficients estimés par les modèles multiniveaux.

¹ On a également inclus dans l'ACP une question dichotomique portant sur la préférence entre l'égalité et la liberté. Le libellé est le suivant : « Laquelle de ces deux opinions est la plus proche de la vôtre ? A : Je trouve que la liberté et l'égalité sont également importantes. Mais s'il fallait choisir l'une ou l'autre, je considérerais que la liberté est plus importante, c'est-à-dire que chacun puisse vivre en liberté et se développer sans contrainte. B : Certainement la liberté et l'égalité sont importantes. Mais s'il fallait que je choisisse, je considérerais que l'égalité est plus importante, c'est-à-dire que personne ne soit défavorisé et que la différence entre les classes sociales ne soit pas aussi forte ».

Figure 1. Demande d'Etat et demande d'égalité (analyse en composantes principales)

	Support for the state	Support for equality
The state should take more responsibility to ensure that everyone is provided for	.70	.05
Government ownership of business and industry should be increased	.64	.26
Competition is harmful	.64	.19
People who are unemployed should have the right to refuse a job they do not want	.60	-.42
Incomes should be made more equal	.02	.76
Equality is more important than freedom	.23	.59

Premières étapes de la modélisation multiniveaux : modèle vide, modèle à constantes aléatoires, modèle à constantes et pentes aléatoires

L'analyse multiniveaux est généralement conduite en plusieurs étapes (Bressoux, 2008²). La première étape consiste à estimer un modèle sans aucune variable explicative (**figure 2. modèle 1**). Ce modèle s'apparente à une ANOVA à effets aléatoires. Il compare le pourcentage de variance de la variable dépendante expliqué par le niveau individuel, et le pourcentage de variance expliqué par le niveau agrégé. Ce modèle dit « vide » permet en fait de savoir si le recours à la modélisation multiniveau est pertinent, en examinant comment la variance de la variable dépendante se distribue entre les différents niveaux d'analyse. La part de variance expliquée par chaque niveau se calcule très simplement en rapportant sa variance à la variance totale. On peut voir ici que 97 % des variations de la demande d'Etat sont à attribuer aux individus, alors que seulement 3 % sont liées aux pays.

Ces valeurs sont également données par le coefficient de corrélation intraclasse. L'ICC représente la part de variance inter-groupes dans la variance totale : il mesure donc le degré de similarité des individus qui appartiennent à une même macro-unité. Ici, un ICC faible signifie que les 23 pays de l'UE se ressemblent beaucoup du point de vue de leur degré d'interventionnisme. Ce résultat indique que la demande d'Etat dépend davantage des différences entre les Européens que des différences entre les pays de l'UE. Bien que faible, la variance de niveau 2 est significative. Il existe donc des différences significatives de soutien à l'Etat d'un pays à l'autre. L'importance de la valeur de l'ICC, la significativité des variances et le test LR attestent ici de la pertinence d'une modélisation multiniveaux et invitent à passer à la seconde étape de l'analyse.

Cette seconde étape (modèle à constantes aléatoires) passe par l'introduction de variables individuelles fixes (**figure 2. modèles 2 et 3**). Les effets fixes correspondent aux paramètres standard d'une régression linéaire. Les modèles 2 et 3 visent ici à évaluer l'impact sur la demande d'Etat du sexe, de l'âge, du niveau de revenus du ménage, du statut d'emploi, de l'orientation politique et de l'intérêt pour la politique. L'analyse étant conduite sur un jeu de données agrégeant plusieurs enquêtes, des variables binaires correspondant aux vagues d'enquête ont été estimées. Ces variables permettent de contrôler l'effet du contexte sociopolitique et socioéconomique lié à chacune des enquêtes. On note que le fait d'être en 2008 ou en 1999 plutôt qu'en 1990 multiplie

² Pascal Bressoux, *Modélisation statistique appliquée aux sciences sociales*, De Boeck, 2008.

significativement la probabilité de développer des attentes fortes à l'égard de l'Etat. Toutes les variables indépendantes étant tenues constantes, les Européens sont donc plus interventionnistes en 1999 qu'en 1990, et plus interventionnistes en 2008 qu'en 1990.

Il est également possible de comparer la part de variance expliquée par chaque modèle relativement au modèle vide. Cela revient à calculer un pseudo-R2 selon une formule de type : (variance estimée par modèle vide - variance estimée par modèle avec X) / variance estimée par modèle vide. L'introduction des variables individuelles permet ici d'expliquer respectivement 4 et 8 % des différences entre individus, et 13 et 25 % des différences entre pays. Les modèles 2 et 3 sont donc plus performants que le modèle vide, même si le gain explicatif est faible (ce qui est assez courant dans la littérature sur les attitudes par rapport à l'Etat). Au total, les modèles 2 et 3 rendent compte d'une portion modeste, mais non négligeable pour ce type d'analyse, des différences entre individus et entre pays.

Figure 2. Modèles multiniveaux estimant la demande d'Etat selon différentes caractéristiques de niveau 1

	Model 1		Model 2		Model 3	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Women			.026***	(.001)	.025***	(.001)
Age (std)			-.001	(.001)	.000	(.001)
Age completed educ. (std)			-.009***	(.001)	-.009***	(.001)
Income - - (ref. ++)			.046***	(.002)	.041***	(.002)
Income -			.038***	(.002)	.035***	(.002)
Income +			.023***	(.002)	.021***	(.002)
Income DKNA			.015***	(.002)	.014***	(.002)
Worker (ref. Self-employed)			.044***	(.003)	.039***	(.003)
Unemployed			.092***	(.004)	.084***	(.004)
Retired or non-worker			.043***	(.003)	.039***	(.003)
Employment status DKNA			.045***	(.006)	.040***	(.006)
EVS 1999 (ref. EVS 1990)			.028***	(.002)	.023***	(.002)
EVS 2008			.043***	(.002)	.041***	(.002)
Left (ref. Right)					.098***	(.002)
Center					.045***	(.002)
Political orientation DKNA					.051***	(.002)
Political interest - (ref. - -)					-.005*	(.002)
Political interest +					-.007***	(.002)
Political interest ++					-.007***	(.002)
Political interest DKNA					.001***	(.003)
Intercept	.398***	(.0059)	.289***	(.006)	.254***	(.006)
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Residual variance (individual-level)	.0258	(.0001)	.0247	(.0001)	.0237	(.0001)
Random intercept variance (country-level)	.0008	(.0002)	.0007	(.0002)	.0006	(.0002)
N individual-level)	65856		65856		65856	
N (country-level)	23		23		23	
Intraclass correlation (%)	3.001		2.62		2.51	
Deviance	-53784.75		-56637.15		-59356.74	
AIC	-53778.75		-56605.15		-59310.73	
BIC	-53751.46		-56459.63		-59101.54	

std=standardized; * p<.05, **p<.01, *** p<.001

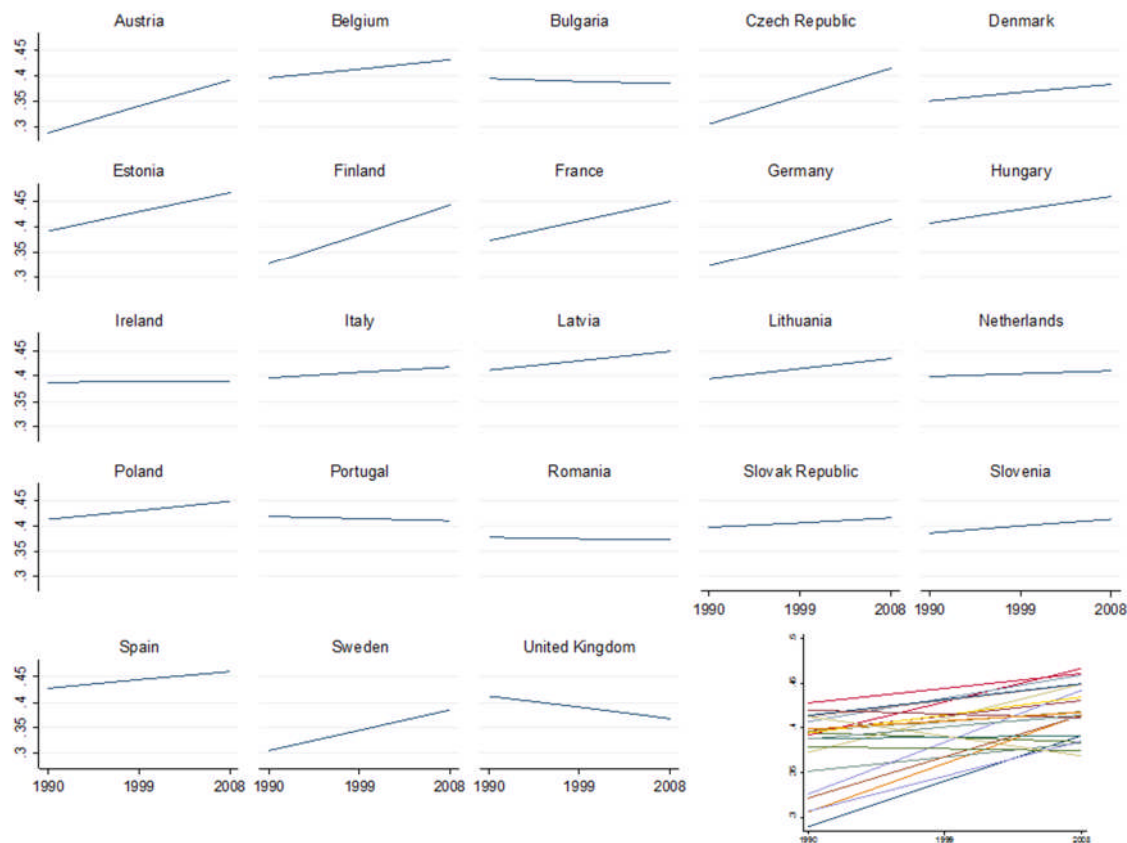
Les modèles 2 et 3 sont des modèles à constantes aléatoires : la demande d'Etat peut prendre des valeurs différentes d'un pays à un autre, mais les droites qui la représentent sont « contraintes » et restent parallèles. Il est possible de faire un pas supplémentaire dans la

modélisation en estimant les droites de régression avec constantes *et* pentes aléatoires. On parle alors de modèle multiniveau complet.

On a graphé les valeurs prédites pour les droites de régression par pays (**figure 3**), en laissant ainsi les coefficients des pentes des droites de régression varier d'un pays à l'autre en fonction de la vague d'enquête. En d'autres termes, on a cherché à tester une hypothèse différentielle en autorisant la relation entre la demande d'Etat et les pays à varier selon le contexte de l'enquête. On a pris au sérieux le fait que la demande d'Etat peut avoir un sens différent d'un pays à un autre, mais aussi à un moment ou à un autre dans les pays de l'UE.

La partie à effet aléatoire de ce modèle donne à voir un autre résultat intéressant. La covariance entre les pentes et les constantes est très significative, avec une corrélation de $-0,9065$. Cela veut dire que les variations entre les niveaux moyens de soutien à l'Etat selon les pays, sont fortement et négativement liés avec les variations entre les niveaux moyens de soutien à l'Etat selon les vagues d'enquête. En clair, les différences entre pays tendent à se réduire d'une vague d'enquête à l'autre. Ce qui indique une forme de convergence des pays de l'UE vers une demande croissante d'intervention étatique, comme le montre bien le graphique global présenté en bas à droite de la figure 3, où l'on voit l'éventail des droites de régression se resserrer entre 1990 et 2008.

Figure 3. Evolution de la demande d'Etat par pays



Introduction de variables contextuelles fixes et modèles avec effets d'interaction entre niveaux

Une étape importante dans la modélisation multiniveau consiste à estimer les effets fixes de variables documentées au niveau agrégé. Il s'agit alors de voir dans quelle mesure ces variables expliquent la variance entre agrégats, et si leur pouvoir prédictif est plus important que les variables indépendantes renseignées au niveau individuel. Comparer les variations et les évolutions au niveau micro et au niveau macrosocial est un enjeu important pour notre analyse.

Les attitudes à l'égard de l'Etat sont-elles par exemple liées au niveau d'inégalités des différents pays de l'Union européenne ? Dans quelle mesure dépendent-elles de l'intervention effective de l'Etat en matière économique et sociale ?

Répondre à ces questions implique de disposer d'indicateurs adaptés pour les deux niveaux d'analyse. La difficulté, ici, est double. D'une part, des données d'opinion documentées tous les neuf ans fournissent une mesure somme toute assez fruste pour appréhender l'évolution de la demande d'Etat. D'autre part, les macro-indicateurs appropriés pour approcher le degré d'inégalités et le degré d'interventionnisme des sociétés européennes restent très mal documentés sur la longue durée. On utilise ici un indicateur classique : le coefficient de Gini (source Eurostat-SILC) pour évaluer le poids des inégalités socioéconomiques (**figure 4. modèle 6**). Un autre macro-indicateur : les dépenses de protection sociale par habitant en 2008 (source Eurostat) permet d'apprécier les différences nationales en termes de régulation étatique (**figure 4. modèle 7**).

Figure 4. Modèles multiniveaux estimant la demande d'Etat selon différentes caractéristiques de niveau 1 et de niveau 2

	Model 6		Model 7	
	Estimate	Std. Err.	Estimate	Std. Err.
Women	.025***	(.001)	.025***	(.001)
Age (std)	.001	(.001)	.001	(.001)
Age completed educ. (std)	-.009***	(.001)	-.009***	(.001)
Income - - (ref. ++)	.046***	(.002)	.046***	(.002)
Income -	.035***	(.002)	.035***	(.002)
Income +	.020***	(.002)	.021***	(.002)
Income DKNA	.017***	(.002)	.017***	(.002)
Worker (ref. Self-employed)	.041***	(.003)	.041***	(.003)
Unemployed	.086***	(.004)	.086***	(.004)
Retired or non-worker	.040***	(.003)	.040***	(.003)
Employment status DKNA	.043***	(.006)	.043***	(.006)
EVS 1999 (ref. EVS 1990)	.024***	(.005)	.024***	(.005)
EVS 2008	.043***	(.010)	.043***	(.010)
Left (ref. Right)	.098***	(.002)	.098***	(.002)
Center	.046***	(.002)	.046***	(.002)
Political orientation DKNA	.055***	(.002)	.055***	(.002)
Political interest - (ref. - -)	-.004*	(.002)	-.004*	(.002)
Political interest +	-.007***	(.002)	-.007***	(.002)
Political interest ++	-.006***	(.003)	-.006***	(.003)
GINI 2008 (std)	.009*	(.004)		
GINI evolution 1999-2008 (std)	-.013**	(.004)		
Social spending 2008 (std)			-.009*	(.005)
Social spending evolution 2000-2008 (std)			.011*	(.005)
GINI evolution x Social spending evolution (std)			.009*	(.005)
Intercept	.248***	(.008)	.251***	(.008)
	Estimate	Std. Err.	Estimate	Std. Err.
Residual variance (individual-level)	.0234	(.0001)	.0234	(.0001)
Random intercept variance (country-level)	.0027	(.0008)	.0032	(.0010)
Variance slope (EVS wave)	.0005	(.0001)	.0005	(.0001)
Covariance random intercept/slope (EVS wave)	-.0011	(.0003)	-.0012	(.0004)
N individual-level)	65856		65856	
N (country-level)	23		23	
Intraclass correlation coefficient (%)	2.15		2.15	
Deviance	-60194.81		-60193.44	
AIC	-60140.81		-60137.44	
BIC	-59895.24		-59882.77	

std=standardized; * p<.05, ** p<.01, *** p<.001

Si l'on compare la part de variance expliquée par chacun de ces deux modèles relativement au modèle vide, on constate que l'introduction des variables contextuelles permet d'expliquer une partie très importante des différences entre individus, mais que le gain explicatif est négatif en ce qui concerne les différences entre pays. Autrement dit, l'introduction des macro-indicateurs fait augmenter la variance des constantes au niveau agrégé. Ce qui se traduit par un gain négatif (en termes de variance expliquée) par rapport aux modèles précédents. Bien que le principe d'un gain explicatif négatif contrarie l'idée qu'on peut avoir d'un modèle explicatif, ce résultat est courant. Comme le souligne Pascal Bressoux (2008), une augmentation de la variance au niveau 2 « est le plus susceptible de se produire quand la plus grande partie, voire la quasi-totalité de la variance résiduelle se répartit sur l'un des niveaux impliqués dans l'analyse. En ce cas, l'introduction d'une ou plusieurs variables explicatives peut conduire à réduire la variance résiduelle du niveau où la variance résiduelle est importante mais à augmenter la variance résiduelle du niveau où la variance résiduelle est faible ». C'est très exactement le cas ici. La quasi-totalité de la variance résiduelle est située, on l'a vu, au niveau des individus. L'introduction des macro-indicateurs fait donc nettement augmenter la variance au niveau des pays (0,0032 dans le modèle 7), alors que celle-ci était faible au départ (0,0008 dans le modèle vide).

Une dernière étape de l'analyse multiniveaux est d'introduire des variables croisées pour tester des interactions entre variables individuelles et variables contextuelles. On a estimé ici un effet d'interaction entre le quartiles des individus les plus défavorisés économiquement (Income--) et le coefficient GINI en statique (Gini 2008 – variable standardisée) et en dynamique (évolution entre 1990 et 2008 – variable standardisée). Si l'on représente graphiquement les valeurs prédites par le modèle (**figure 5**), on constate que les Européens les plus défavorisés suivent la tendance générale : ils sont d'autant plus favorables à l'Etat qu'ils vivent dans un pays où le Gini est élevé, mais ils ne le sont pas davantage que le reste de la population (même si leur demande d'Etat est proportionnellement plus importante que la demande d'Etat « moyenne » des Européens). De même, ils sont d'autant moins favorables à l'Etat qu'ils vivent dans un pays où le Gini a augmenté entre 1990 et 2008, mais ils ne le sont pas moins que la moyenne des Européens.

Figure 5. Variations de la demande d'Etat selon le niveau d'inégalités de revenus des pays

