Université Paris Sud XI

Projet Bi-disciplinaire :

« Peut-on dresser un profil type des électeurs d'Obama en 2008 ? »

Thibault LACHARME
Maxime UCHAN

Projet encadré par M. Jean-Noël Senne et Mme. Marie-Anne Poursat

Table des matières

Introduction

La sphère politique française actuelle est en pleine évolution, les élections présidentielles sont arrivées à leur apogée. Ce sont, pour notre part, les premières que nous vivons à la fois en tant qu'acteurs et en tant que spectateurs. Il nous apparaissait donc comme une évidence de traiter d'un sujet étroitement corrélé avec le contexte récent afin d'en faire un léger parallèle avec notre travail. Cette omniprésence politique dans la vie quotidienne a débuté avec les élections du 45è président des Etats-Unis, Donald Trump et n'a pas cessé jusqu'au duel entre M. Le Pen et E. Macron. Par conséquent, traiter des élections présidentielles des USA nous permettait deux choses : la compréhension du système électoral américain (qui n'est pas l'objet de notre devoir) et la prise de connaissance de l'électorat d'un des présidents les plus emblématiques des USA, Barack Obama.
En effet, notre sujet porte sur l'élection d'Obama en 2008, une élection assez particulière car, rappelons-le, Obama est le premier président noir des Etats Unis. Ce dernier symbolisait d'une part l'espoir d'une communauté souvent sous-représentée et d'autre part le renouveau d'un pays qui s'est dégradé suite aux deux mandats de Georges W. Bush. En effet, la politique menée par ce dernier a divisé le pays sur de nombreux sujets sociaux, géopolitiques ou encore d'ordre moral. Parmi les exemples les plus probants nous pouvons citer :

- l'entrée en guerre avec l'Afghanistan et l'Irak, allant ainsi à l'encontre des préconisations de l'ONU
- l'ouverture de la prison de Guantanamo, dans laquelle ne sont pas respectés les accords de Genève
- la non ratification du protocole de Kyoto symbolisant l'absence d'intérêt pour les problématiques écologiques
- la crise économique des subprimes en fin de second mandat qui est la conséquence d'une politique économique très libérale sous sa tutelle.

Un tel clivage entre les projets portés par ces deux hommes politiques crée une atmosphère toute particulière pour cette élection. Suite aux « échecs » des mesures citées ci-dessus, les américains voient leurs convictions bousculées. Un réel questionnement s'opère du côté de l'électorat -habituellement- Républicain et instaure une critique de celui-ci pâtissant de l'image laissée par Georges W. Bush. Ainsi, le sujet de notre projet bi-disciplinaire nous pousse à étudier les caractéristiques des électeurs d'Obama afin, si possible, d'en dresser un « archétype » dans le but de confronter les idées préconçues que l'on peut avoir sur le sujet à la réalité.

Nous pouvons donc mettre ici en pratique nos modestes connaissances acquises cette année en économétrie et statistique dans le but de définir des caractéristiques particulières correspondants à l'électorat potentiel d'Obama en 2008.

Nous basons ici nos recherches sur une base de données de 3737 individus ayant répondu à une enquête téléphonique de la NAES (National Annenberg Election Survey) sur des questions bien précises peu avant les élections. Nous prenons ici comme variable expliquée la variable indicatrice d'intention de vote pour le candidat Obama (y = 1 si l'individu pense voter Obama, 0 sinon) et allons tenter de dresser un modèle de régression linéaire avec des variables, pour la plupart qualitatives, touchant à diverses caractéristiques pouvant influencer le vote. Enfin nous effectuerons la même opération avec cette fois ci un modèle de régression logistique, plus adapté à la variable endogène étudiée.

Nous commencerons par sélectionner les 13 variables parmi celles présentes dans la base de données (nous n'avons gardé que les plus intéressantes à nos yeux et les plus significatives pour le modèle). Nous effectuerons d'abord une analyse descriptive de celles-ci pour ensuite présenter les deux modèles de régression linéaire et logistique. Nous finirons par une interprétation des résultats ainsi qu'une comparaison des deux modèles.

SELECTION DES VARIABLES

Variable expliquée

Notre variable expliquée (y) est l'intention de vote pour Obama qui est une variable qualitative. En effet trois réponses sont possibles : yes, no et don't know. Il nous faut donc modifier notre variable afin de la transformer en variable binaire. Ainsi, étant donné que nous souhaitons, *in fine*, dresser un profil type de l'électeur d'Obama, nous considérons que « don't know » est identique à « no » car une personne indécise ne représente pas un « électeur type ». Par conséquent, lorsque y=1 cela signifie que l'intention de vote de la personne interrogée est pour Obama (yes), sinon, y=0 fait référence à une intention de vote pour un autre candidat ou une personne indécise (no + don't know). Nous avons donc transformé le modèle initial en un modèle de probabilité linéaire. En effet, une régression avec la méthode des MCO nous simulera une probabilité d'intention de vote pour Obama en fonction des variables explicatives (⇔ P(y=1)).

Cette variable « vote » est recueillie dans un premier temps auprès des personnes présentes dans l'échantillon aléatoire. Cela nous donne une valeur de « vote » empirique. Parmi les 3737 personnes interrogées, seules 2895 y ont répondu, ce qui limite déjà l'échantillon à 2895 observations. Après modification de la variable « vote », nous obtenons une répartition de 50/50 entre « yes » et « no ».

Variables indépendantes/explicatives sélectionnées

Il est important de noter que, pour chacune de ces variables, deux types de réponses un peu particulières sont présentes : « don't know » et « no answer ». Ici « ne sait pas » correspond à une personne n'ayant pas désiré répondre à la question et « pas de réponse » à un problème extérieur ayant pu amener à une absence de réponse (coupure téléphonique avant la fin du sondage…). Ces deux catégories nous permettront de faire des suppositions quant aux raisons de l'absence de réponse et ainsi de suggérer un profil de ces personnes.

WA01_c : sex. Il s'agit là d'une variable qualitative binaire, prenant la valeur 1 si l'individu est un homme et 0 s'il s'agit d'une femme. Nous ne notons aucune valeur manquante et la répartition des deux types d'individus à travers l'échantillon est bonne (43.97 % d'hommes et 56.03 % de femmes). Cette variable est assez significative économiquement dans la régression, il nous semblait très important de la prendre en compte si nous voulions définir un profil type. Nous avons choisi de renommer cette dernière « man » dans notre régression.

Ageslices : modification de la variable WA01_c (age) en tranche d'âges. La variable d'origine étant une variable continue donnant l'âge de chacun des individus ayant répondu, nous avons décidé de la transformer en variable qualitative ordinale pour être plus précis dans notre analyse et pouvoir réellement savoir à quelle tranche d'âge appartient un électeur du candidat Obama. Pour cela nous avons défini quatre tranches d'âge : de 18 à 30 ans, de 31 à 45 ans, de 46 à 65 ans et les plus de 66 ans. Ceci nous permettra donc de faire un parallèle avec la situation professionnelle de chaque individu (à la retraite, étudiant, ...) et déterminer avec plus d'objectivité les motivations de chacun. La variable « age » ne dispose d'aucune valeur manquante et la répartition entre les individus est plutôt équilibrée (ceux ayant entre 30 et 75 ans sont un peu plus représentés que le reste mais il s'agit là d'un éventail d'individus très large). Cette variable sera renommée ageslices à l'avenir.

WA03_c : education. Les individus se sont vus poser la question suivante «What is the last grade or class you completed in school?". Il s'agit donc d'une variable qualitative ordinale regroupant différents diplômes allant du brevet des collèges au doctorat. Aucune valeur manquante n'est à noter et la population est plus ou moins bien répartie mais reste cependant réaliste (les individus ayant un diplôme de fin de lycée, une licence ou un master sont plus représentés, ce qui reste plausible). Nous avons choisi de la modifier pour regrouper certains niveaux d'éducation ensembles afin de disposer de plus grandes significativités économiques. En effet, nous avons réuni sous forme de nouvelle catégorie les personnes « 1. disposant au maximum d'un diplôme de fin de lycée sans avoir continué vers les études supérieures» « 2. Ayant suivi une formation professionnelle post lycée ou des études supérieures sans décrocher de diplôme » « 3. Ayant un diplôme de 2 ou 4 ans après la sortie du lycée (college) » « 4. Ayant poursuivi des études après l'obtention de leurs diplômes universitaire post college » -étant donné que le système scolaire et universitaire américain est différent du nôtre il est parfois compliqué d'établir des équivalences parfaites-. Nous avons pour cela généré une nouvelle variable « educ » issue des groupes de la variable « WA03_c » que nous avions préalablement renommée « education » par soucis de simplicité.

WA04_c : household income. La question suivante est posée : "Last year, what was the total income before taxes of all the people living in your house or apartment?". Cette variable nous fournit le niveau de revenus des ménages par tranche, allant de moins de 10000 $ à 150000 $ ou plus par an. Il s'agit donc d'une variable qualitative ordinale, pour laquelle nous notons 1488 valeurs manquantes et une répartition relativement équilibrée dès que nous dépassons un revenu de 50000$. Malgré les valeurs manquantes et le biais de sélection que peut entraîner la prise en compte de cette variable, celle-ci nous semble tout de même indispensable car c'est un critère social et économique qui révèle beaucoup d'information. D'autant plus que si le revenu n'a aucun impact sur l'intention de vote il sera important de le préciser. Nous la renommerons par la suite HHincome. Pour la même raison que la variable education, nous générons une variable « income » qui se traduit par : « 1. Les ménages gagnant moins de 35,000$/an », « 2. Les ménages gagnant entre 35,000$/an et 75,000$/an » et «3. Les ménages gagnant plus de 75,000$/an ».

WC03_c : race. Les individus devaient répondre à une question simple : « What is your race ? », les choix de réponse regroupant blancs/blancs hispaniques, afro américain/africain hispaniques (black), asiatiques etc. Il s'agit donc d'une variable qualitative nominale, il n'y a aucune valeur manquante pour cette dernière et la population n'est pas réellement bien répartie mais ce qui est assez représentatif des USA (87,37 % de blancs et 6,64 % de noirs). Cette variable parait indispensable à notre régression dans la mesure où nous voulons confronter nos idées préconçues sur l'individu correspondant à l'électeur d'Obama aux résultats de cette étude. Nous l'avons renommé « race ». Nous avons recodé la variable afin que la population noire soit en référence lors du passage sous la forme i.race, nous expliquerons pourquoi dans la partie sur la régression. Dans l'optique d'obtenir des résultats plus significatifs, nous avons regroupé toutes les minorités hormis les noirs dans une catégorie « others ».

WD02_c : religious affiliation. Les individus ont donc dû dire quelle était leur appartenance religieuse, regroupant protestants, catholiques, autres types de chrétien, juifs et toute autre forme de religion (en notant que la religion musulmane est comprise dans la catégorie « other religions »). Nous avons donc affaire à une variable qualitative nominale, aucune valeur manquante et une sur représentation des protestants et catholiques comparé aux autres religions, chose qui semble bien correspondre à la réalité. Cette variable sera renommée religion.

RD01_c : candidate voted for in 2004. La question est simple : pour qui ont voté les individus interrogés aux présidentielles de 2004 ? Nous avons donc Bush, Kerry, Nader ainsi que le vote blanc ou l'abstention. Nous sommes donc face encore une fois à une variable qualitative nominale, ne disposant d'aucune valeur manquante et correspondant bien à la répartition des votes des présidentielles de 2004 (avec en particulier des votes allant pour Bush ou Kerry, candidats des 2 grands partis américains). Il paraît logique de penser que cette variable sera fortement significative dans notre modèle, ce qui nous a poussé à la prendre en considération. Nous l'avons renommée candidate2004. Nous recodons afin que Kerry soit la réponse de référence.

ABo05_c : Obama is a strong leader. Il était demandé aux individus de donner sur une échelle allant de 0 à 10 leur ressenti sur la phrase « Obama est un bon leader », allant de « does not apply at all» à « applies extremely well». Il s'agit là d'une variable qualitative ordinale pour laquelle seulement 41 valeurs sont manquantes et les extrêmes sont plus représentés (0, 8 ou 10), raison pour laquelle nous avons choisi de regrouper certaines réponses dans une même variable « lead ». Dans un premier temps nous avons inversé toutes les valeurs symétriquement afin qu'en 0, ce soit « applies extremely well » jusqu'à 10 qui devient « does not apply at all ». Encore une fois cela sert à mettre en référence le groupe le plus favorable à une élection d'Obama. Ensuite, nous considérons trois groupes pour simplifier : ceux qui pensent qu'Obama est un réel leader, ceux qui sont plutôt indifférents, ceux qui ne le pensent pas. Pour cela nous générons une variable « lead » définie par : « 0. leader=0+1+2 », « 5. leader=3+4+5+6+7 », « 10. leader=8+9+10 ». Cette mesure nous permet de limiter les imprécisions liées au jugement incertain que les personnes peuvent apporter à cette question.

ABo12_c : Obama shares my values. Même type de variable que la précédente, sur une échelle de 0 à 10 dire à quel point la phrase « Obama partage mes valeurs » semble juste. Variable qualitative ordinale, 41 valeurs sont manquantes et le même constat s'opère sur les réponses, seules 0, 5, 8 et 10 ressortent ce qui nous a poussé à former des groupes. Avant tout nous avons renommé en « values ». Exactement comme pour la variable précédente, nous avons réordonner symétriquement « values » puis généré « value » qui se code exactement comme « lead » s'est codée en fonction de « leader ».

CDb01_c : US should withdraw or keep troops in Iraq. Les individus se sont vus présenter plusieurs politiques relatives à la position militaire des Etats Unis en Irak et devaient désigner laquelle leur paraissait la plus adaptée. Les choix regroupent donc le retrait des troupes immédiat, leur retrait progressif, leur maintien tant que le gouvernement n'est pas stable ou aucune de ces politiques. Cette variable est donc une variable qualitative nominale, aucune valeur manquante n'est à noter et les réponses sont relativement bien réparties, chose qui semble plausible étant donné qu'il s'agissait d'une question divisant encore beaucoup le pays à l'époque. Cette variable semble tout à fait intéressante pour notre modèle dans la mesure où Obama mettait en avant dans son programme qu'il souhaitait retirer les troupes américaines d'Irak, laissant ainsi la responsabilité aux autorités irakiennes qu'il considérait maintenant comme aptes à gérer le pays. Nous avons renommé cette variable « withdrawtroops » dans notre modèle. Encore une fois nous prenons en référence la position sur le sujet du sondé la plus similaire à celle d'Obama, soit « withdraw as soon as possible ».

CEa01_c : Abortion should be available or restricted. Pour cette variable, les individus devaient dire quelle position sur l'avortement leur semblait le plus en accord avec leur point de vue. Les propositions sont les suivantes : autorisée quel que soit la situation, autorisée avec certaines limites, interdite sauf dans des cas extrêmes de viol ou d'inceste, toujours interdite ou aucune de ces propositions. C'est donc une variable qualitative nominale pour laquelle aucune valeur n'est manquante et la question semble toujours diviser le peuple américain (33.64 % sont pour l'autorisation quel que soit la situation et 31.47% seulement pour les cas extrêmes). Connaissant la position d'Obama sur la question, il est intéressant de garder cette variable pour bien comprendre qui sont ses électeurs. Nous la renommerons « abortion » pour notre modèle. Position de référence : « avalaible to anyone » bien que la position d'Obama soit légèrement différente sur le sujet, il semble davantage en faveur de l'avortement sous réserve de certaines conditions. La prise en référence de « avalaible to anyone » ne change pas les résultats attendus par rapport à « available with stricter limits ».

CFb01_c : Protecting environment or growing economy should be more important. Les individus doivent signaler ici quelle affirmation semble la plus proche de leurs convictions sur la question. Nous avons donc : l'environnement doit être la priorité, la croissance économique doit être mise en avant, les deux de manière équivalente, aucune des deux. Il d'agit donc d'une variable qualitative nominale ne disposant d'aucune variable manquante et étant assez représentative de l'incertitude des citoyens américains face à cette problématique (56.06 % préfèrent la croissance économique et 36.69 % l'environnement). Nous la renommons « environment ». Nous prendrons en référence « environment should be top priority ».

MA01_c : Party ID. Les individus doivent ici simplement dire à quel parti ils s'identifient le plus. Les choix sont donc : républicains, démocrates, indépendants et autres. Cette variable fait donc partie des qualitatives nominales, avec 0 valeurs manquantes et une bonne représentation de l'électorat américain (27.80 % pour le parti républicain, 36.31 % pour le parti démocrate et 30.37 % pour le parti indépendant américain).Tout comme candidate2004, cette variable est hautement significative pour notre modèle, il était donc impensable de ne pas la prendre en considération dans la mesure où elle permettrait aussi de déterminer les raisons du choix pour certains républicains de finalement voter pour Obama. Cette dernière sera renommée « partyaffiliation ». Evidemment, nous prenons en référence les personnes s'identifiant comme « démocrates ».

La régression linéaire multiple (RLM)

D'après les variables que nous avons présentées, notre régression linéaire multiple est :

```
. reg vote man i.ageslices i.educ i.income i.race i.religion i.candidate2004 i.lead i.value i.withdrawtroops i.abo
> rtion i.environment i.partyaffiliation
```

Source	SS	df	MS
Model	330.766675	57	5.80292413
Residual	146.054791	1852	.078863278
Total	476.821466	1909	.249775519

Number of obs	1910
F(57, 1852)	73.58
Prob > F	0.0000
R-squared	0.6937
Adj R-squared	0.6843
Root MSE	.28083

(*) nous ne pouvons mettre une capture d'écran de l'ensemble étant donné la taille du tableau

Le modèle que nous avons estimé à l'aide de la méthode des MCO est le suivant :

$vote = \beta_0 + \beta_1 man + (\beta_2 ageslices_1 + \ldots + \beta_4 ageslices_3) + (\beta_5 educ_2 + \ldots + \beta_9 educ_{999}) + (\beta_{10} income_5 + \ldots + \beta_{12} income_{999}) + (\beta_{13} race_1 + \ldots + \beta_{16} race_{999}) + (\beta_{17} religion_2 + \ldots + \beta_{24} religion_{999}) + (\beta_{25} candidate2004_1 + \ldots + \beta_{30} candidate2004_{999}) + (\beta_{31} lead_5 + \ldots + \beta_{34} lead_{999}) + (\beta_{35} value_5 + \ldots + \beta_{38} value_{999}) + (\beta_{39} withdrawtroops_2 + \ldots + \beta_{43} withdrawtroops_{999}) + (\beta_{44} abortion_2 + \ldots + \beta_{49} abortion_{999}) + (\beta_{50} environment_2 + \ldots + \beta_{54} environment_{999}) + (\beta_{55} partyaffiliation_1 + \ldots + \beta_{58} partyaffiliation_{999}) + u$

Interprétation du modèle

Ci-dessous la première partie de notre régression

vote	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
man	.0038514	.0135047	0.29	0.776	-.0226347	.0303375
ageslices						
1	-.0277096	.0330168	-0.84	0.401	-.0924638	.0370445
2	-.010054	.0315866	-0.32	0.750	-.0720031	.0518952
3	.0000265	.0336842	0.00	0.999	-.0660365	.0660895
educ						
2	.0091793	.0207956	0.44	0.659	-.031606	.0499646
3	.0236853	.0195331	1.21	0.225	-.0146239	.0619945
4	.0379816	.0211684	1.79	0.073	-.0035348	.0794981
998	-.3622911	.2036509	-1.78	0.075	-.7617006	.0371184
999	-.0340094	.31735	-0.11	0.915	-.6564107	.5883919
income						
5	.0101136	.0204454	0.49	0.621	-.0299848	.0502121
7	.0015498	.0213725	0.07	0.942	-.0403668	.0434665
10	.0051239	.0308195	0.17	0.868	-.0553208	.0655686

N.B. : afin de ne pas le répéter à chaque interprétation, nous considérons que l'interprétation des $\boldsymbol{\beta}_k$, $k \in \{1, \ldots, 58\}$, se fait toutes choses égales par ailleurs, i.e. tous autres facteurs fixés.

β_1 (man) = 0.0038, cela signifie que d'après le modèle, en moyenne, la probabilité qu'un homme vote Obama augmente de 0.38 point de pourcentage par rapport à une femme (qui est le groupe de référence). On remarque que la significativité économique est bien faible, cela démontre que le sexe n'a que peu d'incidence sur l'intention de vote.

β_2 ($ageslices_1$) = -0.0277. Ainsi d'après le modèle précédent, en moyenne la probabilité qu'une personne ayant entre 31 et 45 ans vote pour Obama diminue de 2.77 points de pourcentages par rapport aux 18-30 ans. Encore une fois, la significativité économique est peu élevée. Nous n'interpréterons pas tous les paramètres car l'analyse est similaire pour chaque sous-groupe. On comprend donc que l'âge semble peu représentatif de l'intention de vote.

β_7 ($education_4$) = 0.038. Par conséquent le modèle nous indique qu'en moyenne détenir le niveau de diplôme le plus élevé augmente la probabilité de vote de 3.8 points de pourcentage par rapport aux personnes sans diplôme universitaire. Lorsque l'on regarde bien, on voit que plus le niveau d'éducation augmente, plus le différentiel de vote augmente, on suppose donc une relation croissante entre l'intention de vote pour Obama et le niveau d'éducation.

β_{10} ($income_5$) = 0.0101. Cette valeur nous apprend qu'en moyenne, une personne gagnant entre 35,000$ et 75,000$ par an a une probabilité supérieure de 1.01 points de pourcentage de voter pour Obama par rapport à une personne gagnant moins de 35,000$/an. En regardant les autres coefficients, on observe que la classe « moyenne » a une probabilité de vote supérieure pour Obama que les autres (en considérant que la classe moyenne est celle que nous venons de traiter). Deux remarques s'imposent. Dans un premier temps, la classe sociale n'est pas uniquement conditionnée par le salaire et de plus notre tranche de salaire est très large. Dans un second temps, les significativités économiques sont faibles ce qui nous pousserait davantage à penser que le salaire du ménage n'a pas beaucoup d'influence sur le vote futur.

ci-dessous la seconde partie de la régression :

race						
1	-.0995761	.0271673	-3.67	0.000	-.1528578	-.0462945
3	-.116466	.0411686	-2.83	0.005	-.1972078	-.0357241
998	-.1248253	.149068	-0.84	0.402	-.4171843	.1675337
999	-.0756296	.0789214	-0.96	0.338	-.2304139	.0791546
religion						
2	.0338343	.0170805	1.98	0.048	.0003352	.0673334
3	.0057467	.0296349	0.19	0.846	-.0523746	.0638679
4	.0012909	.0382431	0.03	0.973	-.0737133	.0762951
5	.0313834	.0390228	0.80	0.421	-.04515	.1079168
6	.0474084	.0185138	2.56	0.011	.0110983	.0837185
7	.0815778	.0537827	1.52	0.129	-.0239034	.1870589
998	.0094936	.0875726	0.11	0.914	-.1622579	.181245
999	.1095984	.132256	0.83	0.407	-.1497881	.368985
candida~2004						
1	-.2716146	.0219017	-12.40	0.000	-.3145693	-.2286599
3	.1157504	.0695906	1.66	0.096	-.020734	.2522347
4	-.1288413	.0619863	-2.08	0.038	-.2504116	-.0072709
5	-.0957401	.0280982	-3.41	0.001	-.1508475	-.0406327
998	-.0982932	.0654809	-1.50	0.134	-.2267173	.0301309
999	-.4499205	.1699187	-2.65	0.008	-.7831729	-.1166681
lead						
5	-.1486962	.0197467	-7.53	0.000	-.1874243	-.1099681
10	-.200817	.0284917	-7.05	0.000	-.2566962	-.1449379
998	.0350769	.0776917	0.45	0.652	-.1172957	.1874495

β_{13} ($race_1$) = -0.0996. Donc, selon le modèle ci-dessus, en moyenne, la probabilité de voter pour Obama lorsqu'on est de race (origine) « white / white hispanic » codée 1, est diminuée de 9.96 points de pourcentage si l'on compare avec une personne de race « afro-american/ black hispanic ». Nous avions cette intuition dès le départ car Obama pouvait, à l'époque, devenir le premier président noir des USA. Nous nous doutions qu'un tel symbole serait soutenu par la communauté noire des Etats-Unis. En revanche, plus surprenant, il semblerait que les autres minorités (indian american, asian, …), que nous avons codées ensemble en 3, soient moins favorable à l'élection d'Obama que les personnes de race « blanche » (-0.116<-0.0996). Malgré tout, cette différence reste légère et peut être la conséquence d'une erreur de précision car l'échantillon du groupe « blanc » est 18 fois supérieur à l'échantillon de « other races ».

β_{17} ($religion_2$) = 0.0338. Le modèle prédit qu'en moyenne, être catholique augmente la probabilité de vote pour Obama de 3.38 points de pourcentage par rapport à être protestant. Cela peut paraître étonnant car la religion d'Obama est protestante, cependant cette valeur n'est pas très significative économiquement. Il reste à noter d'intéressant que la probabilité de vote pour Obama augmente de 3.14 points de pourcentages (respectivement de 4.74 points de pourcentage) lorsque l'individu est de religion type « others » (respectivement « no denomination » i.e. croyant mais sans religion) dont font partie les musulmans (toujours comparé aux protestants), communauté historiquement moins appréciée par le parti républicain, plus conservateur concernant la religion. On peut donc aussi supposer que B. Obama apporte un projet qui touche la communauté « sans religion », nous essaierons d'en explorer la raison plus tard.

β_{25} ($candidate2004_1$) = -0.2716. Cette variable que l'on a supposé très importante préalablement, confirme notre intuition selon laquelle l'historique de vote joue un rôle majeur dans l'intention de vote. Ce coefficient démontre que selon le modèle précédent, en moyenne, une personne ayant voté Bush (candidat républicain) en 2004 a une probabilité de vote pour Obama qui diminue de 27.16 points de pourcentage par rapport à une personne ayant voté Kerry (candidat démocrate). Concrètement, cela signifie que les individus ne changent pas facilement de convictions et ce malgré l'impopularité et le bilan négatif de Georges Bush. Autre coefficient remarquable, celui devant $candidate2004_3$ (Nader). Nader était le candidat du parti écologique des USA, on remarque que les personnes ayant voté Nader auront, en moyenne selon le modèle, plus tendance à voter Obama en 2008 que les personne ayant voté Kerry. Cela peut être dû à deux choses : le fait qu'Obama prenne en compte à part entière l'écologie dans son programme mais aussi le fait que le vote de 2008 pouvait être contestataire à l'encontre de Bush et des républicains.

β_{32} ($lead_{10}$)=-0.200817. Dans la mesure où nous avons pris comme groupe de référence une personne considérant Obama comme un bon leader, il paraît donc normal qu'un individu ayant une vision diamétralement opposée sur le sujet soit moins enclin à voter pour ce dernier. Ici nous avons une différence de 20.0817 points de pourcentage entre ces deux profils, ce qui est très significatif économiquement. Cette variable, contrairement aux précédentes, correspond plus au ressenti de l'individu interrogé sur le candidat Obama. Il parait donc intéressant de noter que l'image du candidat auprès des électeurs a un réel impact sur leur choix final, voire même plus que l'appartenance à une certaine catégorie socio-économique. Nous verrons notamment cette hypothèse se confirmer avec les variables qui suivent étant plus en rapport avec l'idée que les électeurs se font du candidat et leur avis sur certaines problématiques.

value						
5	−.2395518	.0209978	−11.41	0.000	−.2807337	−.1983699
10	−.3261454	.0301926	−10.80	0.000	−.3853605	−.2669302
998	−.1360956	.0713404	−1.91	0.057	−.2760117	.0038205
withdrawtr~s						
2	−.0568381	.0177006	−3.21	0.001	−.0915533	−.0221228
3	−.1217715	.021522	−5.66	0.000	−.1639815	−.0795615
4	−.1295895	.0492519	−2.63	0.009	−.2261846	−.0329944
998	−.2636175	.0810823	−3.25	0.001	−.4226399	−.1045952
999	−.2436897	.1324351	−1.84	0.066	−.5034274	.016048
abortion						
2	−.044857	.0184346	−2.43	0.015	−.0810118	−.0087023
3	−.0375769	.0189397	−1.98	0.047	−.0747223	−.0004314
4	−.0189853	.0255969	−0.74	0.458	−.0691871	.0312164
5	.0726661	.0480679	1.51	0.131	−.0216068	.1669391
998	.0745095	.0875886	0.85	0.395	−.0972732	.2462922
999	−.0167872	.0658974	−0.25	0.799	−.1460282	.1124539
environment						
2	−.0400469	.0155872	−2.57	0.010	−.0706172	−.0094766
3	−.0371319	.0353677	−1.05	0.294	−.1064967	.032233
4	−.0232913	.0486448	−0.48	0.632	−.1186957	.072113
998	.0070505	.0580385	0.12	0.903	−.1067772	.1208782
999	−.036073	.1312961	−0.27	0.784	−.2935769	.2214308
partyaffil~n						
1	−.1680104	.0235387	−7.14	0.000	−.2141755	−.1218453
3	−.0872286	.0186008	−4.69	0.000	−.1237093	−.0507478
4	−.1651993	.0480149	−3.44	0.001	−.2593683	−.0710303
998	−.1568441	.0553582	−2.83	0.005	−.2654151	−.0482731
999	−.0136993	.081011	−0.17	0.866	−.1725817	.1451832
_cons	1.129776	.0458352	24.65	0.000	1.039882	1.21967

β_{37}(value$_{10}$)=--0.3261454. Par rapport à des individus considérant qu'Obama partage leurs valeurs, un électeur pensant complètement différemment verra sa probabilité de vote pour Obama diminuer de 32.61454 points de pourcentage. Cette variable est donc elle aussi hautement significative pour déterminer le choix de vote lors de ces élections. Il s'agit là d'un résultat plausible dans la mesure où les électeurs seront amenés à voter pour le candidat les représentant le plus, à la fois dans les mesures économiques qu'il met en place et les valeurs qu'il défend.

β_{41}(withdrawtroops$_3$)= -0.1218. Nous comparons donc ici deux groupes ayant un avis divergent sur la politique de retrait des troupes en Irak, représentant l'une des mesures du précédent président la plus sujette à des critiques de la part du peuple américain. Le groupe de référence souhaite ici retirer les troupes immédiatement tandis que le groupe étudié préfèrerait maintenir les forces armées jusqu'à mise en place d'un gouvernement stable. Il y a donc entre ces deux groupes une différence de (-)12.18 points de pourcentage sur la probabilité de vote pour Obama, cette variable est donc elle aussi significative pour notre modèle. Dans la mesure où Obama propose dans son programme de retirer les troupes il paraît normal que les personnes favorables au maintien de ces dernières aient une probabilité plus faible de vote pour le candidat démocrate.

β_{46} (abortion$_4$) = -0.019. L'interprétation est la suivante : en moyenne, selon le modèle, un individu qui pense que l'avortement ne doit être autorisé sous aucun prétexte a une probabilité de vote pour Obama plus faible de 1.9 points de pourcentage par rapport à un individu qui pense que l'avortement doit être très accessible. La position d'Obama sur le sujet est un peu différente : il considère que l'avortement doit être utilisé avec parcimonie, c'est-à-dire que son avis se rapproche davantage de la catégorie abortion$_2$. Malgré la valeur de β_{44}, cela reste logique car les personnes très décomplexées sur le sujet de l'avortement se rapprocheront du candidat ayant l'avis le plus similaire au leur, soit celui d'Obama (par rapport à celui de Mc Cain, le candidat républicain en 2008).

β_{50} (environment$_2$) = -0.04. Ainsi, selon le modèle, en moyenne, les gens qui favoriseraient l'expansion de l'économie aux dépens de l'écologie, auraient une probabilité de voter pour Obama plus faible de 4 points de pourcentage par rapport aux personnes favorisant l'écologie. Cela reste cohérent vis-à-vis de la position d'Obama sur le sujet qui a considéré que l'écologie faisait partie intégrante de son projet en tant que président (en passant par la protection des espaces naturels protégés notamment).

β_{55} (partyaffiliation) = -0.168. Variable très significative économiquement. Cette valeur n'est pas surprenante car elle relève de l'historique politique de l'individu que nous avons précédemment considéré comme étant peu variable. Une personne se considérant comme proche du parti républicain verra sa probabilité de voter pour Obama diminuer de 16.8 points de pourcentage par rapport à un individu affilié au parti démocrate. Nous pourrions penser que cette variable est fortement liée à candidate2004 cependant ,après vérification, l'absence d'une de ces deux variables a bien un impact négatif sur le modèle et amène à la surestimation de l'effet de celle gardée. En effet, il est possible de penser qu'une personne étant auparavant affiliée au parti républicain puisse avoir une intention de vote différente suite aux résultats mitigés des deux mandats de Georges W. Bush.

β_0 (constante) = 1.1298.
Le modèle prédit qu'en moyenne, une femme protestante noire (ou afro américaine ou hispanique noire de peau) qui :

- A entre 18 et 30 ans
- N'a pas de diplôme universitaire
- Gagne moins de 35,000$/year
- A voté pour Kerry en 2004
- Pense qu'Obama est un excellent leader
- Partage les valeurs d'Obama
- Souhaite que les USA retirent leur troupe d'Irak le plus rapidement possible
- Pense que l'avortement doit être accessible sans conditions sur la grossesse
- Estime que l'environnement doit être une priorité
- Se sent proche du parti démocrate

Aura une probabilité de vote en faveur d'Obama estimée à 1.1298.

Bien évidemment, cela n'a pas de sens car une probabilité ne peut excéder 1 tout comme elle ne peut être inférieure à 0. Ceci nous amène à discuter des limites rencontrées par notre modèle de régression linéaire.

Significativité des variables et du modèle

Nous pouvons commencer par noter que le modèle est globalement très significatif statistiquement. En effet, nous avons ici une F-stat de 73.58 et une p-valeur pour le modèle très proche de 0.

Pour ce qui est des coefficients, nous en avons 23 sur 40 significatives au seuil de 10% (p valeur), 21 à 5% et 15 à 1%, sachant que nous ne prenons pas en compte les coefficients associés à « don't know » et « no answer » (respectivement 998 et 999). Il y a donc un peu plus de la moitié de coefficients significatifs au seuil de 10%, ce qui est relativement peu. Cependant nous pouvons tenter de comprendre les raisons de cette faible significativité statistique en portant notre attention sur certains coefficients du modèle.

Prenons par exemple les coefficients associés à income, nous pouvons voir qu'aucun n'est significatif statistiquement. Nous avons pu noter lors de l'étude des différentes variables que cette dernière avait 1488 variables manquantes. En nous referrant au calcul de la T-stat, celle-ci dépend à la fois de la valeur du coefficient (non estimée) et de sa standard error. Or nous savons que la variance estimée des coefficients dépend positivement de l'écart type estimé et négativement à la fois de la taille de l'échantillon et de la multicolinéarité entre les variables explicatives. Dans notre cas, l'échantillon est réduit à 2249 observations sur le revenu du ménage. De plus, nous savons, à l'issu des travaux de Mincer, que le niveau d'éducation et le salaire d'un individu sont fortement corrélés. Malgré le fait que l'income porte sur le ménage, d'après le principe de l'endogamie sociale, au sein d'un couple on retrouve des niveaux d'éducation respectivement très similaires, ce qui induit *in fine* de la multicolinéarité. Un raisonnement similaire peut s'appliquer aux autres paramètres peu significatifs.

Enfin, nous avons dans ce modèle un R^2 ajusté de 0.6843, cela signifie donc que nous parvenons à expliquer 68.43 % de la variance de la variable dépendante avec ce modèle, ce qui est élevé.

Limite du MPL

Nous avons jusqu'à présent utilisé une régression linéaire, ce qui nous donne un modèle de probabilité linéaire « MPL ». Estimer un tel modèle à l'aide de la méthode des MCO n'est pas une mauvaise idée car son utilisation ne biaise pas l'interprétation des résultats. Seulement, il faut que le modèle soit saturé i.e. les probabilités estimées doivent être comprises dans [0,1]. Ce n'est pas le cas ici, il nous suffit de regarder la constante. Ce que nous préconisons est de changer de modèle et de passer à un modèle de régression logistique que l'on appelle modèle logit.

N.B. : Un bref rappel sur le modèle logistique est présent dans les annexes, il porte sur l'interprétation des résultats majoritairement

Le modèle de régression logistique

```
. logit vote man i.ageslices i.educ i.income i.race i.religion i.candidate2004 i.lead i.value i.withdrawtroops i.abortio
> n i.environment i.partyaffiliation

Logistic regression                             Number of obs     =       1906
                                                LR chi2(55)       =    1790.31
                                                Prob > chi2       =     0.0000
Log likelihood = -424.69784                     Pseudo R2         =     0.6782
```

vote	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
man	.1323568	.1887654	0.70	0.483	-.2376165	.5023302
ageslices						
1	-.3116072	.4399847	-0.71	0.479	-1.173961	.550747
2	.0153722	.4231418	0.04	0.971	-.8139704	.8447149
3	.1285541	.455646	0.28	0.778	-.7644957	1.021604
educ						
2	.0325727	.2872354	0.11	0.910	-.5303982	.5955437
3	.179054	.2722247	0.66	0.511	-.3544967	.7126047
4	.4576042	.2994566	1.53	0.126	-.1293199	1.044528
998	-3.040085	1.601316	-1.90	0.058	-6.178607	.0984367
999	(empty)					
income						
5	.1277321	.2726016	0.47	0.639	-.4065572	.6620213
7	-.1065416	.2917443	-0.37	0.715	-.6783499	.4652667
10	.320861	.4616847	0.69	0.487	-.5840243	1.225746
race						
1	-1.700477	.5231715	-3.25	0.001	-2.725874	-.6750798
3	-2.057413	.6637885	-3.10	0.002	-3.358414	-.7564113
998	-2.728169	2.23667	-1.22	0.223	-7.111962	1.655624
999	-.9116361	1.294118	-0.70	0.481	-3.44806	1.624788
religion						
2	.4188721	.2399841	1.75	0.081	-.0514881	.8892323
3	.1873173	.4549965	0.41	0.681	-.7044595	1.079094
4	-.0245435	.4997608	-0.05	0.961	-1.004057	.9549696
5	-.0746535	.534188	-0.14	0.889	-1.121643	.9723357
6	.4031046	.2503099	1.61	0.107	-.0874938	.893703
7	1.210933	.754105	1.61	0.108	-.2670857	2.688952
998	.3065431	1.433054	0.21	0.831	-2.50219	3.115277
999	1.948	4.347887	0.45	0.654	-6.573702	10.4697

(Cf. Annexe pour la régression complète)

Pour ne pas que notre dossier soit purement descriptif, nous n'interprèterons pas tous les coefficients mais nous le ferons pour quelques uns. L'interprétation des autres paramètres se fait de manière similaire par rapport à leur groupe de référence. Tâchons d'interpréter les coefficients que nous jugeons les plus intéressants, dont nous justifierons la sélection. Encore une fois nous raisonnons à autres facteurs fixés, i.e. toutes autres choses égales par ailleurs.

β_{13} (race) = -1.700. D'après le modèle de régression logistique ci-dessus, en moyenne, le log odds (expliqué en annexe) est de -1.7. Cela signifie que la probabilité de vote pour Obama sachant que l'individu est de « race » blanche est bien inférieure à la probabilité de vote pour Obama lorsque l'individu est de « race » noire. Enfin, en passant à l'exponentielle, on observe que ce ratio de probabilité est de 0.1827 (odds). En passant à l'inverse, on obtient que la probabilité de vote pour Obama d'une personne noire et 5.47 fois supérieure à celle d'une personne blanche. On observe ici une différence davantage significative économiquement par rapport à la régression linéaire multiple, ce qui justifie la sélection de ce coefficient pour l'interprétation. Ceci conforte notre idée selon laquelle l'identité et l'appartenance à une communauté (ici à travers la couleur de peau) influence l'intention de vote envers Obama.

β_3 ($ageslices_2$) = 0.0154. On remarque que le signe a changé par rapport à la RLM. Le log odds augmente de 0.0154 lorsque l'individu appartient aux 46-65 ans par rapport aux individus appartenant aux 18-30 ans (en moyenne, selon le modèle). Bien évidemment cette différence est très mince, ce qui nous indique que la significativité économique dans le modèle de régression linéaire multiple l'était aussi malgré le changement de signe. Finalement, ce qui ressort de la contradiction mise en évidence, c'est que la probabilité de vote pour Obama n'est pas (ou peu) influencée directement par l'âge du citoyen.

β_0 (constante) = 5.7483
Ainsi, lorsque toutes les variables sont égales à 0, nous obtenons une probabilité :

$$P(Y = 1) = \frac{1}{(1+e^{-5,7483})} = 0.996822$$

Ceci est la probabilité de vote pour Obama sous tous les critères de β_0 (cf interprétation de la constante dans la RLM).

Avec ces deux interprétations, nous avons l'étendue des modifications faites par le modèle de régression logistique quant à l'interprétation des résultats. Ainsi soit une amplification de l'effet initial se remarque (cas de « race ») ce qui conforte l'idée selon laquelle le modèle de régression linéaire multiple ne nous induit pas en erreur sur le sens des observations. Soit, d'autre part, lorsque le paramètre change de signe, nous observons que les paramètres initiaux étaient très faibles et donc le passage en logit change le signe en laissant de faibles coefficients, ce qui nous laisse suggérer que l'effet de base était quasi nul.

Significativité

Nous pouvons voir que le pseudo R^2 du modèle de régression logistique est de 0.6782, donc assez proche de la valeur obtenue dans notre précédent modèle.

La significativité globale se mesurant avec un test de chi-2 à 55 degrés de liberté est à 1790 et une p-valeur de 0.0000 ce qui nous montre que notre modèle logistique est très significatif.

Quant aux paramètres, nous voyons que 18 sur 40 sont significatifs au seuil de 10%, 16 le sont au seuil à 5% et 12 à 1%. Encore une fois, très peu de paramètres semblent être significatifs statistiquement.

Il semblerait qu'ici l'utilisation du modèle logit nous apporte seulement des résultats plus réalistes par rapport au modèle de régression linéaire. Effectivement, les R^2 ajustés sont similaires, il en est de même pour les significativités globales. En revanche, le modèle logit apporte moins de significativités individuelles dans les paramètres.

Pour répondre à cette question, nous avons fait différentes régressions en enlevant certains paramètres. Ce que nous donne trois cas différents.

Voici notre régression initiale.

```
. reg vote man i.ageslices i.educ i.income i.race i.religion i.candidate2004 i.lead i.value i.withdrawtroops i.abortion i.environm
> ent i.partyaffiliation
```

Source	SS	df	MS		
Model	330.766675	57	5.80292413	Number of obs =	1910
Residual	146.054791	1852	.078863278	F(57, 1852) =	73.58
				Prob > F =	0.0000
Total	476.821466	1909	.249775519	R-squared =	0.6937
				Adj R-squared =	0.6843
				Root MSE =	.28083

<u>Cas n-°1</u> :

Nous enlevons de notre régression linéaire multiple « income » et « educ ».

(Cf. Annexe pour la régression complète)

```
. reg vote man i.ageslices i.race i.religion i.candidate2004 i.lead i.value i.withdrawtroops i.abortion i.environment i.partyaffil
> iation
```

Source	SS	df	MS		
Model	503.638246	51	9.87525974	Number of obs =	2879
Residual	216.017189	2827	.076412165	F(51, 2827) =	129.24
				Prob > F =	0.0000
Total	719.655436	2878	.250054008	R-squared =	0.6998
				Adj R-squared =	0.6944
				Root MSE =	.27643

Nous pouvons voir que le R^2 ajusté a augmenté, passant de 0.6843 à 0.6944, tout en ne remarquant pas de grandes différences au niveau de la significativité économique de chacun des paramètres. Ceci pourrait donc nous faire penser que ces deux variables ne sont pas « utiles » pour notre modèle. Il nous paraissait cependant important de prendre en compte ces deux variables dans notre modèle si nous voulions dresser un profil type, afin de déterminer l'influence de ces dernières sur le choix final de l'individu.

Au vu de ces résultats nous pouvons donc considérer que le niveau de salaire et d'éducation n'ont finalement que peu d'impact sur l'intention de vote d'un individu, invalidant ainsi l'intuition que nous pouvions avoir sur le sujet. Nous pouvons néanmoins rappeler qu'il ne s'agit que de notre modèle, or ce dernier n'est pas forcément proche de la réalité malgré les résultats encourageants que nous avons obtenus et ce pour plusieurs raisons, chose que nous traiterons dans les limites et critiques du modèle.

Cas n-°2 :

Prenons maintenant la régression de base et retirons les variables liées au ressenti des individus sur le candidat Obama (soient lead, value, withdrawtroops, abortion et environment), nous obtenons :

(Cf. Annexe pour la régression complète)

```
. reg vote man i.ageslices i.educ i.income i.race i.religion i.candidate2004 i.partyaffiliation
```

Source	SS	df	MS
Model	274.231868	35	7.83519624
Residual	205.246528	1885	.1088841
Total	479.478397	1920	.249728332

Number of obs	1921
F(35, 1885)	71.96
Prob > F	0.0000
R-squared	0.5719
Adj R-squared	0.5640
Root MSE	.32998

Nous notons tout d'abord une nette différence au niveau du R^2 ajusté du modèle, passant de 0.6843 à 0.5640. Les variables que nous avons retiré du modèle semblent donc réellement importantes pour la qualité de ce dernier. Nous pouvons aussi remarquer que la significativité économique de certaines variables a tendance à augmenter (la variable race par exemple pour laquelle le coefficient associé au fait d'être de race blanche passe de -0.09 à -0.18), nous pouvons donc en déduire que l'effet des variables concernées est surestimé lorsque nous omettons les variables en question.

Ces deux régressions nous auront donc permis de comparer l'influence de variables traitant plutôt du milieu social auquel appartiennent les individus à celle des variables prenant en compte le ressenti des personnes sur le candidat. Nous en avons donc conclu avec ce modèle que l'image qu'ont les électeurs du candidat, étant quelque chose de très personnel dans la mesure où il s'agit d'un jugement de valeur, joue un rôle plus important que les critères socio-économiques des individus dans notre modèle.

Cas n-°3 :

Nous réitérons l'expérience en omettant « candidate2004 » et « partyaffiliation ».

(Cf. Annexe pour la régression complète)

```
. reg vote man i.ageslices i.income i.educ i.race i.religion i.lead i.value i.withdrawtroops i.abortion i.environment
```

Source	SS	df	MS
Model	303.192178	46	6.5911343
Residual	173.629288	1863	.093198759
Total	476.821466	1909	.249775519

Number of obs	1910
F(46, 1863)	70.72
Prob > F	0.0000
R-squared	0.6359
Adj R-squared	0.6269
Root MSE	.30528

De même que précédemment, le R^2 diminue, d'environ 7 points de pourcentages. Cela nous indique déjà que l'historique de vote et l'appartenance à une idéologie prennent une place non négligeable dans l'explication de notre variable dépendante. De plus, de nombreuses significativités économiques ont encore augmentée par rapport au modèle initial. L'exemple le plus marquant étant sur la variable « value » qui capte une grande partie de l'effet omis par les deux variables « partyaffiliation » et « candidate2004 ». En réalité, la variable « value » prend en compte de nombreux facteurs omis dans son estimation car partager les valeurs d'un candidat se retranscrit par le partage de ses idées et de son idéologie.

Nous pouvons donc suggérer d'après notre modèle, que les variables liées au ressenti que l'on porte sur un candidat sont essentielles dans la détermination du vote. Vient ensuite les variables faisant référence à l'historique de vote et d'appartenance à une idéologie, ce qui illustre le fait que les individus ne changent pas facilement d'opinion. Enfin les variables du type socio-économique sont moins (voire pas) représentative de l'intention de vote.

Jusqu'à présent, nous avons parlé d'une seule limite à laquelle nous nous sommes confrontée : le biais d'endogénéité relatif à la forme fonctionnelle. Pour pallier ce problème nous avons introduit le modèle de régression logistique.

Nous savons que notre modèle est imparfait et il est primordial de comprendre les autres sources de ces imperfections. Nous traiterons ces autres types de biais et tenterons d'apporter des préconisations dans notre dernière partie portant sur les limites.

Limites

Suite aux cours que nous avons suivi en économétrie, il ressort deux types de biais que nous maîtrisons : le biais de sélection et le biais d'endogénéité. Commençons par élucider le biais de sélection.

Le biais de sélection est relatif à l'échantillonnage aléatoire. Ainsi, avoir du biais de sélection signifie que l'on ne dispose pas d'un échantillon représentatif de la population américaine. Il s'avère que malgré le fait que l'organisme chargé de cette enquête tente de minimiser le biais de sélection, il persiste. En effet, cette enquête étant téléphonique, il faut que les ménages disposent d'un téléphone, et bien que la grande majorité de la population en dispose, il se peut que dans les foyers les plus modestes il n'y ait pas de téléphone fixe. De plus, il faut que l'individu appelé soit disponible, par conséquent une personne n'étant que peu à domicile, car sa situation professionnelle ne le permet pas, ne pourra pas répondre à l'enquête. Cette indisponibilité liée au travail se retrouve chez les personnes exerçant un métier à forte charge de travail et fortes responsabilités majoritairement.

Enfin, dans le cadre de certaines variables, à l'image de « income » nous décelons de nombreuses variables manquantes ainsi qu'une représentation limitée des ménages les plus modestes. Ceci n'est pas un aléa, aux US, ce sont les ménages les plus pauvres qui, peut-être par « honte », ne souhaitent pas transmettre leurs revenus. (Annotation : en France ce biais s'applique aussi aux ménages les plus riches)

D'autre part, le second type de biais (endogénéité) se traduit par quatre sources : la forme fonctionnelle (déjà traitée), l'erreur de mesure, la causalité inverse et l'omission de variables.

Tout d'abord, étudions le cas où une erreur de mesure lors de l'enquête viendrait biaiser la régression. Nous avons pu remarquer, au cours de l'étude des variables, que de nombreuses questions cherchaient à capter l'image que les individus interrogés pouvaient avoir du candidat. Le jugement que pouvaient avoir ces derniers était donc très subjectif, et il est difficile d'en obtenir une mesure très précise. En effet les personnes interrogées se sont vus demander de noter sur une échelle de 0 à 10 à quel point l'affirmation présentée leur paraissait juste, pouvant aller

de « Obama est un bon leader » à « Obama a assez d'expérience pour être président ». Nous pouvons donc penser que certaines réponses ont pu être données rapidement et sans grande réflexion de la part de l'interrogé. Nous avons deux de ces variables dans notre modèle, à savoir « lead » et « value », il paraît donc indispensable de prendre en compte ce biais comme l'une des raisons expliquant l'imperfection de notre modèle. Néanmoins, nous avons tenté de pallier ce problème en regroupant les réponses en trois groupes différents faisant ressortir les notes les plus proches de 0, celles les plus proches de 5 et celles étant les plus proches de 10. Ceci dans le but, dans un premier temps, d'augmenter la taille des échantillons autour des valeurs les plus représentées.

Concernant la causalité inverse, l'intention de vote pour Obama peut avoir un impact sur toutes les variables recensant les avis comme « value » ou « lead ». En effet, une personne qui souhaite déjà voter pour Obama avant l'enquête aura probablement tendance à conforter son choix en surestimant les « notes » données pour « value » et « lead » par exemple. Ce procédé d'auto persuasion pourrait s'affilier à de la dissonance cognitive. Ainsi l'intention de vote pourrait influencer certaines variables indépendantes.

Enfin, il est évident que le biais de variable omise est omniprésent dans notre modèle, mais il faut savoir que ce biais peut être minimisé mais jamais anéanti (du moins avec les modestes connaissances que nous avons). Déterminer les variables omises revient à comprendre la composition du résidu. Pour commencer, dans notre base de données, nous avions de nombreuses variables qui pouvaient être très intéressantes au niveau des critères socio-économiques. Par exemple, l'orientation sexuelle, peut être un critère intéressant car aucune de nos variables ne seraient corrélées directement avec cette dernière. Néanmoins, cette variable avait beaucoup trop de valeurs manquantes. Nous disposions aussi de variables en relation avec la présence médiatique des candidats, nous avons décidé de ne pas les utiliser par choix. En effet, la fréquence à laquelle les individus regardent la télévision, lisent le journal est une variable utile dans la prédiction du vote. La présence de deux candidats au sein des médias n'est pas équitable en général, à l'image de Trump/Clinton, un candidat est généralement plus mis en avant qu'un autre, ce qui influence le choix des électeurs.

De plus, nous avons songé à d'autres variables omises n'étant pas dans la base de données sélectionnée par M. Senne. Parmi celles-ci nous en avons relevé deux principalement : le secteur d'activité professionnelle et le lieu de résidence de l'individu. Le lieu de résidence se réfèrerait à tous les critères pouvant influencer l'intention de vote comme : la proximité avec une frontière, l'enclavement de l'état de résidence, la concentration d'emploi, …

CONCLUSION

Il est maintenant temps de répondre à notre question initiale qui est, rappelons-le, *« peut-on dresser un profil type des électeurs d'Obama en 2008 ? »*. D'après notre modèle, nous obtenons finalement que l'électeur type d'Obama se distingue selon les caractéristiques suivantes (déterminées avec le modèle de régression logistique) :

- Etant un Homme
- Ayant plus de 65 ans
- Détenant un des diplômes universitaires les plus élevés
- Dont le ménage gagne plus de 75,000$/an
- De « race » afro américaine/noire hispanique/noire
- Etant athée/agnostique
- Ayant voté pour Nader en 2004 (parti écologiste)
- Qui partage les valeurs défendues par Obama
- Qui pense qu'Obama est un Leader
- Souhaitant que les USA retirent immédiatement leurs troupes d'Irak
- Qui ne partage aucun des avis proposés par le questionnaire sur le thème de l'avortement
- Qui ne souhaite ni mettre l'accent en particulier sur l'écologie ni sur l'économie du pays
- Etant affilié au parti démocrate

Néanmoins, nous nous sommes ici uniquement référés aux coefficients de la régression. Ainsi, si nous souhaitons prendre en compte les significativités (économiques et statistiques) pour minimiser le risque de se tromper sur le profil, nous obtenons des résultats différents :

- L'âge, le niveau d'éducation ainsi que le salaire ne sont plus des critères déterminants
- La religion de l'individu a plus de chance d'être catholique
- Son vote en 2004 aurait été en faveur de Kerry
- Sa position sur l'avortement serait en faveur d'un accès assez libre à l'avortement
- L'individu serait enclin à mettre l'écologie au cœur des priorités

Les modifications que nous venons d'apporter prennent en compte le nombre de personnes vérifiant ces critères par rapport au nombre total d'observations, nous voulons donc savoir s'il est bien représentatif de l'échantillon étudié. Ainsi pour des variables telles qu' « abortion », l'échantillon d'individus ayant répondu « none of them » n'était que de 75 sur le nombre total d'observations contrairement aux individus étant favorables à un accès libre à l'avortement en toute circonstance représentant 33.64% des personnes interrogées.

Finalement, que pouvons-nous retenir de notre travail ?

Premièrement, nous avons vu les limites du modèle de régression linéaire multiple.

De plus, nous sommes capables de dire que certaines variables ont une vocation plus prédictive que d'autres, nous l'avons constaté en séparant les régressions selon la « catégorie » des variables : socio-économiques, liées au ressenti, …

Enfin nous avons ressenti à quel point mener à bien un projet économétrique cohérent, sachant que nous ne disposions pas toujours des compétences nécessaires ou encore des données suffisantes, pouvait être difficile.

Bibliographie

Sur la question de l'avortement :

http://www.washingtontimes.com/news/2015/jan/22/obama-deeply-committed-preserving-abortion-rights/

http://www.cnsnews.com/news/article/terence-p-jeffrey/human-rights-chair-obama-abortion-president

Sur la question de l'environnement :

http://www.politifact.com/truth-o-meter/promises/obameter/subjects/environment/

https://www.theguardian.com/environment/climate-consensus-97-per-cent/2016/nov/02/barack-obama-is-the-first-climate-president

Sur la question de la guerre en Irak :

http://www.reuters.com/article/us-usa-politics-obama-idUSN0923153320070212

https://www.nytimes.com/elections/2008/president/issues/iraq.html

Aide à la compréhension du modèle logit :

https://fr.wikipedia.org/wiki/R%C3%A9gression_logistique

http://spss.espaceweb.usherbrooke.ca/pages/stat-inferentielles/regression-logistique/interpretation.php

ANNEXE

Modèle de régression logistique

La régression logistique permet de tester un modèle de régression dont la variable dépendante est dichotomique (qui prend les valeurs 0 ou 1). Une telle variable (Y) suit une loi Binomiale de paramètre (n,p) où n est la taille de l'échantillon et p (0 < p < 1) la probabilité de succès, soit ici la probabilité de vote pour Obama.

Un modèle de régression linéaire multiple est défini comme étant :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

Un modèle de régression logistique se défini comme ceci :

$$\ln\left(\frac{P(Y=1)}{P(Y=0)}\right) = \beta 0 + \beta 1 X1 + \beta 2 X2 + \ldots + \beta k Xk$$

Ceci est le *log odds*. Dans le cas de la comparaison en *log odds,* nous comparons nos valeurs des paramètres obtenus à 0. Si la valeur est négative alors le quotient des probabilités est inférieur à 1 et ainsi que la probabilité d'échec est supérieure à la probabilité de succès. Cette interprétation est celle que nous utilisons au sein de la régression logistique.

Ce qui équivaut à :

$$\frac{P(Y=1)}{P(Y=0)} = \exp\left(\beta 0 + \beta 1 X1 + \beta 2 X2 + \ldots + \beta k Xk\right)$$

Que l'on appelle *l'odds*.

En réajustant, on obtient :

$$\mathrm{P}(\mathrm{Y}=1) = \frac{1}{1 + \exp -(\beta 0 + \beta 1 X1 + \beta 2 X2 + \ldots + \beta k Xk)}$$

Car P(Y = 0) = 1 – P(Y = 1)

Par conséquent, la probabilité P(Y=1) sera comprise entre 0 et 1, ce qui est l'effet recherché dans le passage en logit.

Régression logistique complète

```
. logit vote man i.ageslices i.educ i.income i.race i.religion i.candidate2004 i.lead i.value i.withdrawtroops i.abortio
> n i.environment i.partyaffiliation
```

Logistic regression

Number of obs	1906
LR chi2(55)	1790.31
Prob > chi2	0.0000
Log likelihood = -424.69784	Pseudo R2 = 0.6782

vote	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
man	.1323568	.1887654	0.70	0.483	-.2376165	.5023302
ageslices						
1	-.3116072	.4399847	-0.71	0.479	-1.173961	.550747
2	.0153722	.4231418	0.04	0.971	-.8139704	.8447149
3	.1285541	.455646	0.28	0.778	-.7644957	1.021604
educ						
2	.0325727	.2872354	0.11	0.910	-.5303982	.5955437
3	.179054	.2722247	0.66	0.511	-.3544967	.7126047
4	.4576042	.2994566	1.53	0.126	-.1293199	1.044528
998	-3.040085	1.601316	-1.90	0.058	-6.178607	.0984367
999	(empty)					
income						
5	.1277321	.2726016	0.47	0.639	-.4065572	.6620213
7	-.1065416	.2917443	-0.37	0.715	-.6783499	.4652667
10	.320861	.4616847	0.69	0.487	-.5840243	1.225746
race						
1	-1.700477	.5231715	-3.25	0.001	-2.725874	-.6750798
3	-2.057413	.6637885	-3.10	0.002	-3.358414	-.7564113
998	-2.728169	2.23667	-1.22	0.223	-7.111962	1.655624
999	-.9116361	1.294118	-0.70	0.481	-3.44806	1.624788
religion						
2	.4188721	.2399841	1.75	0.081	-.0514881	.8892323
3	.1873173	.4549965	0.41	0.681	-.7044595	1.079094
4	-.0245435	.4997608	-0.05	0.961	-1.004057	.9549696
5	-.0746535	.534188	-0.14	0.889	-1.121643	.9723357
6	.4031046	.2503099	1.61	0.107	-.0874938	.893703
7	1.210933	.754105	1.61	0.108	-.2670857	2.688952
998	.3065431	1.433054	0.21	0.831	-2.50219	3.115277
999	1.948	4.347887	0.45	0.654	-6.573702	10.4697
candida~2004						
1	-1.621449	.233674	-6.94	0.000	-2.079442	-1.163456
3	1.95915	1.060374	1.85	0.065	-.1191437	4.037444
4	-.4306507	.6573372	-0.66	0.512	-1.719008	.8577066
5	-.4638363	.3309767	-1.40	0.161	-1.112539	.1848662
998	-.7362016	.7410493	-0.99	0.320	-2.188631	.7162283
999	(empty)					
lead						
5	-1.16586	.2189478	-5.32	0.000	-1.59499	-.7367301
10	-2.951661	.669712	-4.41	0.000	-4.264273	-1.63905
998	.6200701	1.051945	0.59	0.556	-1.441705	2.681845
value						
5	-1.527751	.2278709	-6.70	0.000	-1.974369	-1.081132
10	-3.746005	.6268119	-5.98	0.000	-4.974534	-2.517476
998	-1.120674	.8354062	-1.34	0.180	-2.75804	.5166916

withdrawtr~s						
2	-.6103664	.2352939	-2.59	0.009	-1.071534	-.1491989
3	-1.331637	.2764714	-4.82	0.000	-1.873511	-.7897633
4	-1.481928	.6317458	-2.35	0.019	-2.720127	-.2437287
998	-3.009498	1.13912	-2.64	0.008	-5.242132	-.7768633
999	-3.37857	1.725204	-1.96	0.050	-6.759907	.002767
abortion						
2	-.50571	.2374896	-2.13	0.033	-.9711811	-.0402389
3	-.501834	.2513839	-2.00	0.046	-.9945374	-.0091306
4	-.3238745	.3873328	-0.84	0.403	-1.083033	.4352838
5	1.101733	.7591723	1.45	0.147	-.386217	2.589684
998	.7540045	1.090659	0.69	0.489	-1.383648	2.891657
999	-.0298688	.7821976	-0.04	0.970	-1.562948	1.50321
environment						
2	-.4771037	.1964547	-2.43	0.015	-.8621479	-.0920595
3	-.4369117	.4486614	-0.97	0.330	-1.316272	.4424485
4	.034883	.6372684	0.05	0.956	-1.21414	1.283906
998	-.0685554	.752943	-0.09	0.927	-1.544297	1.407186
999	.122309	1.835858	0.07	0.947	-3.475906	3.720524
partyaffil~n						
1	-1.833068	.3123089	-5.87	0.000	-2.445182	-1.220954
3	-.709334	.2170178	-3.27	0.001	-1.134681	-.283987
4	-1.084617	.6662008	-1.63	0.104	-2.390347	.221112
998	-1.594931	.5970547	-2.67	0.008	-2.765137	-.4247252
999	.3481803	1.160346	0.30	0.764	-1.926055	2.622416
_cons	5.748277	.7384177	7.78	0.000	4.301004	7.195549

Regression linéaire multiple sans « income » et « educ »

```
. reg vote man i.ageslices i.race i.religion i.candidate2004 i.lead i.value i.withdrawtroops i.abortion i.environment i.partyaffil
> iation
```

Source	SS	df	MS
Model	503.638246	51	9.87525974
Residual	216.017189	2827	.076412165
Total	719.655436	2878	.250054008

Number of obs	= 2879
F(51, 2827)	= 129.24
Prob > F	= 0.0000
R-squared	= 0.6998
Adj R-squared	= 0.6944
Root MSE	= .27643

vote	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
man	.0078604	.0107025	0.73	0.463	-.0131251	.0288459
ageslices						
1	-.0025286	.0263176	-0.10	0.923	-.0541322	.049075
2	.0021847	.0252162	0.09	0.931	-.0472593	.0516287
3	.0073777	.0265912	0.28	0.781	-.0447624	.0595178
race						
1	-.0908936	.0223136	-4.07	0.000	-.1346461	-.0471411
3	-.0933303	.0323717	-2.88	0.004	-.1568048	-.0298557
998	-.1443891	.1431253	-1.01	0.313	-.4250297	.1362516
999	-.1446313	.0624333	-2.32	0.021	-.2670507	-.022212
religion						
2	.0309996	.01342	2.31	0.021	.0046856	.0573136
3	.0020731	.0240653	0.09	0.931	-.0451143	.0492605
4	.0093497	.0319318	0.29	0.770	-.0532624	.0719617
5	.057586	.0310833	1.85	0.064	-.0033622	.1185343
6	.0322708	.0150312	2.15	0.032	.0027976	.0617441
7	.1081443	.0379112	2.85	0.004	.033808	.1824807
998	-.0346815	.0705566	-0.49	0.623	-.1730293	.1036662
999	.1414756	.0967404	1.46	0.144	-.0482134	.3311646
candida~2004						
1	-.2489416	.0173976	-14.31	0.000	-.2830549	-.2148284
3	.0428983	.0523889	0.82	0.413	-.0598261	.1456228
4	-.1202615	.0536224	-2.24	0.025	-.2254044	-.0151185
5	-.0857411	.0214504	-4.00	0.000	-.1278012	-.0436811
998	-.0909967	.0538745	-1.69	0.091	-.196634	.0146406
999	-.224426	.1267985	-1.77	0.077	-.4730528	.0242009
lead						
5	-.1734063	.0157326	-11.02	0.000	-.2042549	-.1425577
10	-.2212388	.0228916	-9.66	0.000	-.2661246	-.1763529
998	.0129325	.0633444	0.20	0.838	-.1112735	.1371384
999	-.123199	.2269594	-0.54	0.587	-.5682218	.3218238
value						
5	-.23039	.016793	-13.72	0.000	-.2633179	-.1974622
10	-.3274411	.0239887	-13.65	0.000	-.3744781	-.280404
998	-.1334713	.0549478	-2.43	0.015	-.241213	-.0257296
999	-.3227215	.2269229	-1.42	0.155	-.7676728	.1222298
withdrawtr~s						
2	-.0400567	.0140457	-2.85	0.004	-.0675976	-.0125157
3	-.1376124	.0171736	-8.01	0.000	-.1712864	-.1039384
4	-.0997763	.0386029	-2.58	0.010	-.175469	-.0240837
998	-.2818279	.058458	-4.82	0.000	-.3964525	-.1672033
999	-.2421366	.1278548	-1.89	0.058	-.4928348	.0085615
abortion						
2	-.0330929	.0148992	-2.22	0.026	-.0623074	-.0038785
3	-.0315961	.0146635	-2.15	0.031	-.0603484	-.0028438
4	-.0093885	.0198487	-0.47	0.636	-.048308	.0295309
5	.0518847	.0398145	1.30	0.193	-.0261836	.1299531
998	.0999889	.0684078	1.46	0.144	-.0341454	.2341232
999	-.0832993	.0562094	-1.48	0.138	-.1935149	.0269163
environment						
2	-.0344694	.0122856	-2.81	0.005	-.058559	-.0103798
3	-.0347232	.0291704	-1.19	0.234	-.0919206	.0224742
4	.0213459	.0392118	0.54	0.586	-.0555406	.0982325
998	.0144301	.0499809	0.29	0.773	-.0835726	.1124327
999	-.023086	.107341	-0.22	0.830	-.2335607	.1873886
partyaffil~n						
1	-.1631044	.0187702	-8.69	0.000	-.1999091	-.1262997
3	-.081424	.0145587	-5.59	0.000	-.1099708	-.0528772
4	-.1350678	.0402744	-3.35	0.001	-.214038	-.0560976
998	-.0788669	.0432587	-1.82	0.068	-.1636887	.005955
999	.0408184	.0597322	0.68	0.494	-.0763046	.1579415
_cons	1.12809	.0336785	33.50	0.000	1.062053	1.194127

Régression linéaire multiple sans « value », « lead », « abortion », « withdrawtroops » et « abortion »

```
. reg vote man i.ageslices i.educ i.income i.race i.religion i.candidate2004 i.partyaffiliation
```

Source	SS	df	MS		
Model	274.231868	35	7.83519624	Number of obs	1921
Residual	205.246528	1885	.1088841	F(35, 1885)	71.96
				Prob > F	0.0000
				R-squared	0.5719
				Adj R-squared	0.5640
Total	479.478397	1920	.249728332	Root MSE	.32998

vote	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
man	-.039747	.0155639	-2.55	0.011	-.0702713	-.0092227
ageslices						
1	-.0355968	.0382901	-0.93	0.353	-.1106922	.0394986
2	-.006292	.0366049	-0.17	0.864	-.0780824	.0654984
3	-.0195098	.0388868	-0.50	0.616	-.0957754	.0567559
educ						
2	.0148747	.0240417	0.62	0.536	-.0322764	.0620258
3	.0467874	.0225355	2.08	0.038	.0025902	.0909845
4	.0793238	.0243163	3.26	0.001	.0316341	.1270134
998	-.2399155	.238756	-1.00	0.315	-.7081694	.2283383
999	.0972056	.3713004	0.26	0.794	-.6309975	.8254087
income						
5	.0136318	.0234547	0.58	0.561	-.0323681	.0596316
7	.0109178	.0244116	0.45	0.655	-.0369589	.0587945
10	.0129363	.0347909	0.37	0.710	-.0552965	.081169
race						
1	-.1805267	.0310606	-5.81	0.000	-.2414434	-.11961
3	-.1933627	.0478105	-4.04	0.000	-.2871298	-.0995957
998	-.2252365	.1740351	-1.29	0.196	-.5665582	.1160851
999	-.0851779	.0911303	-0.93	0.350	-.2639048	.0935489
religion						
2	.0443448	.0197945	2.24	0.025	.0055234	.0831661
3	.013254	.0346497	0.38	0.702	-.0547019	.0812098
4	.0243718	.044476	0.55	0.584	-.0628556	.1115993
5	.0478004	.0444965	1.07	0.283	-.0394672	.135068
6	.0740517	.0212578	3.48	0.001	.0323604	.115743
7	.1478875	.0622953	2.37	0.018	.0257124	.2700625
998	.0701653	.1018415	0.69	0.491	-.1295686	.2698992
999	.168222	.1541562	1.09	0.275	-.1341127	.4705568
candida~2004						
1	-.4888016	.0225935	-21.63	0.000	-.5331125	-.4444907
3	.0987955	.0812653	1.22	0.224	-.060584	.258175
4	-.2520909	.0720008	-3.50	0.000	-.3933005	-.1108813
5	-.1252034	.0319758	-3.92	0.000	-.1879152	-.0624917
998	-.2078788	.0739009	-2.81	0.005	-.3528149	-.0629427
999	-.7411786	.1702061	-4.35	0.000	-1.074991	-.4073665
partyaffil~n						
1	-.3606677	.0258787	-13.94	0.000	-.4114215	-.3099138
3	-.1973559	.0211156	-9.35	0.000	-.2387682	-.1559435
4	-.3216673	.0554056	-5.81	0.000	-.43033	-.2130046
998	-.1902498	.0632058	-3.01	0.003	-.3142105	-.0662892
999	-.2023001	.0900871	-2.25	0.025	-.3789811	-.0256192
_cons	1.010462	.0487059	20.75	0.000	.9149387	1.105985

Régression linéaire multiple sans « candidate2004 » et « partyaffiliation »

```
. reg vote man i.ageslices i.income i.educ i.race i.religion i.lead i.value i.withdrawtroops i.abortion i.environment
```

Source	SS	df	MS		
Model	303.192178	46	6.5911343	Number of obs	= 1910
Residual	173.629288	1863	.093198759	F(46, 1863)	= 70.72
				Prob > F	= 0.0000
Total	476.821466	1909	.249775519	R-squared	= 0.6359
				Adj R-squared	= 0.6269
				Root MSE	= .30528

vote	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
man	.0189668	.0145685	1.30	0.193	-.0096054	.0475391
ageslices						
1	-.0175287	.0353121	-0.50	0.620	-.0867841	.0517268
2	-.0034493	.0334206	-0.10	0.918	-.068995	.0620964
3	.0117667	.0353955	0.33	0.740	-.0576523	.0811857
income						
5	-.0165073	.021952	-0.75	0.452	-.0595604	.0265458
7	-.0440363	.0228556	-1.93	0.054	-.0888615	.000789
10	-.0247377	.0332471	-0.74	0.457	-.0899432	.0404678
educ						
2	.0082561	.0224362	0.37	0.713	-.0357466	.0522588
3	.0121516	.0210155	0.58	0.563	-.0290648	.0533681
4	.0413448	.0227911	1.81	0.070	-.0033539	.0860436
998	-.3135692	.2212122	-1.42	0.157	-.747419	.1202806
999	-.0093604	.3440676	-0.03	0.978	-.6841589	.6654381
race						
1	-.1648778	.0292478	-5.64	0.000	-.2222398	-.1075159
3	-.1739452	.0444763	-3.91	0.000	-.2611737	-.0867167
998	-.1986679	.1614942	-1.23	0.219	-.5153965	.1180606
999	-.195497	.0846657	-2.31	0.021	-.3615466	-.0294474
religion						
2	.0480754	.0185127	2.60	0.009	.0117676	.0843832
3	-.011203	.0321274	-0.35	0.727	-.0742125	.0518065
4	.037925	.0414689	0.91	0.361	-.0434054	.1192554
5	.0388335	.0422925	0.92	0.359	-.0441122	.1217791
6	.0549589	.0200443	2.74	0.006	.0156473	.0942705
7	.0661421	.058223	1.14	0.256	-.048047	.1803313
998	-.0572935	.0934937	-0.61	0.540	-.240657	.12607
999	.1300406	.1431947	0.91	0.364	-.1507983	.4108795
lead						
5	-.1809814	.0213345	-8.48	0.000	-.2228235	-.1391393
10	-.2274122	.0308687	-7.37	0.000	-.287953	-.1668714
998	.0331565	.0841929	0.39	0.694	-.1319659	.1982789
value						
5	-.3311514	.0219655	-15.08	0.000	-.374231	-.2880718
10	-.4822114	.0313077	-15.40	0.000	-.5436132	-.4208095
998	-.2294167	.0761956	-3.01	0.003	-.3788545	-.0799789
withdrawtr~s						
2	-.0842309	.0190498	-4.42	0.000	-.1215921	-.0468697
3	-.2433008	.0221191	-11.00	0.000	-.2866815	-.19992
4	-.1837402	.0532665	-3.45	0.001	-.2882086	-.0792718
998	-.2843893	.0875103	-3.25	0.001	-.4560178	-.1127608
999	-.2639678	.1428217	-1.85	0.065	-.5440751	.0161396
abortion						
2	-.0731221	.0199282	-3.67	0.000	-.1122059	-.0340382
3	-.0981901	.0202186	-4.86	0.000	-.1378437	-.0585365
4	-.1012596	.027282	-3.71	0.000	-.1547661	-.047753
5	.0403517	.0520956	0.77	0.439	-.0618201	.1425235
998	.0167122	.0945135	0.18	0.860	-.1686512	.2020755
999	-.0461128	.0708667	-0.65	0.515	-.1850993	.0928736
environment						
2	-.0897766	.0166352	-5.40	0.000	-.1224022	-.057151
3	-.0761468	.0382101	-1.99	0.046	-.151086	-.0012077
4	-.0892407	.0526184	-1.70	0.090	-.1924378	.0139564
998	-.0099236	.0629424	-0.16	0.875	-.1333686	.1135213
999	-.0751168	.1420997	-0.53	0.597	-.3538081	.2035746
_cons	1.201194	.0473917	25.35	0.000	1.108247	1.29414