ASSESSMENT OF THE INFLUENCE OF EDUCATION LEVEL ON VOTING INTENTION FOR THE EXTREME RIGHT IN FRANCE.

Joseph Rynkiewicz[*,*], Mohamed Raouf Benmakrelouf[**], Wafa Karouche[**]
[*]Université de Paris I, 90 rue de Tolbiac, Paris, France.
[**]Université des sciences et de la technologie Houari Boumediene, Bab Ezzouar, Algiers, Algeria.

ABSTRACT
In France, the Front National has been a growing political party in the last 30 years. After years of stagnant economy, French voters have come to mistrust the political elite, and have been increasingly receptive the Front National straight-talking approach. The most consistent findings in social research on ethnic attitudes is the negative association between educational attainment and ethnic prejudice: People with higher education are less prejudiced toward ethnic out groups than those with lower education. We might expect, then, that high education will generally prevent people from voting for the extreme right, regardless of their position in the labor market. This point of view, however, is not unanimous and there is an intellectual elite within the extreme right, especially among angry academic white males who are uneasy with the gains of feminism and believe in a left-wing media conspiracy. Hence the link between Extreme Right voting and education is not so obvious. The aim of this paper is first to build a model which takes into account the causality structure of variables (Bayesian Network) and secondly to assess the influence of education on the voting intention for extreme right-wing party by taking into consideration all the possible confounding variables.

KEYWORDS: Logistic regression, Confounding variable, Bayesian network

MSC: 62P25

RESUMEN
En Francia, el Frente Nacional ha sido un partido político cada vez más exitoso en los últimos 30 años. Después de años de estancamiento de la economía , los votantes franceses han llegado a desconfiar de la élite política, y han sido cada vez más receptivos al enfoque simplista del Frente Nacional. Los hallazgos más consistentes en la investigación social sobre las actitudes étnicas describen la asociación negativa entre el nivel educativo y el prejuicio étnico : Las personas con educación superior tienen menos prejuicios hacia los grupos étnicos que aquellos con menor educación. Podríamos esperar , entonces, que la educación superior por lo general impide votar por la extrema derecha, independientemente de la posición en el mercado de trabajo. Este punto de vista , sin embargo , no es unánime y hay una élite intelectual dentro de la extrema derecha , sobre todo entre los hombres blancos enojados “academicos” que están incómodos por ejemplo con los logros del feminismo y creen en una conspiración izquierdista. De ahí que la relación entre la votación extrema derecha y la educación no es tan obvia. El objetivo de este trabajo es primero construir un modelo que tiene en cuenta la estructura de causalidad de las variables (Red Bayesiana) y en segundo lugar, evaluar la influencia de la educación en la intención de voto para el partido de extrema derecha , tomando en cuenta todas las posibles variables de confusión.

1. INTRODUCTION

In France, the Front National has been a growing political party in the last 30 years. After years of stagnant economy, French voters no longer trust the political elite or any decisions from the European Union, and have been increasingly receptive to the Front National straight-talking approach. One of the most consistent findings in social research on ethnic attitudes is the negative association between educational attainment and ethnic prejudice: People with higher education are less prejudiced toward ethnic out groups than are those with lower education. This relationship has been established in empirical research over time as well as in

[*]joseph.rynkiewicz@univ-paris1.fr.

different countries (Schuman et al. [10]; Vogt [12]). This point of view, however, is not unanimous as Müller [7] has pointed out, there is an intellectual elite within the extreme right, especially among angry academic white males who are uneasy with the gains of feminism and believe in a left-wing media conspiracy. Hence the link between Extreme Right voting and education is not so obvious. We think that, for this link, a lot of variables may be confounding variables and that the perceived relationship between level of education and voting intention may be mis-estimated due to the failure to account for such confounding factors. A classical example of such phenomenon can be found in the chapter devoted to the Simpson paradox in Agresti [1], moreover it is well known that confounding is a causal concept so we have to take into account most of causal relationships of our model in order to assess the real effect of education level. In this paper, we build a model which takes into consideration the causality structure of variables (Bayesian Network) and we assess the influence of education on the voting intention for extreme right-wing party thanks to the Back-door adjustment (see Pearl [8]). The paper is organized as follows: Section 2 establishes the model from a selection of variables of social profile but also from feelings and beliefs variables. We select the important variables by using information criterion and additive logistic regression. In the next section, employing the selected variables we construct a model with the causality structure using the bnlearn package (see Scutari [11]). Finally, we deduce from the Bayesian network model the influence of the education level on the voting intentions for the extreme right by taking into account all the possible confounding variables.

2. A GENERAL VOTING MODEL

An significant step in the models construction is the selection of variables explaining the voting intention. The model is built using data from a representative survey of the CEVIPOF [5], which were gathered using face-to-face interviews with people aged 18 years and over. It is composed of survey of representative sample of French voters registered on the voting rolls and broken out by region. This survey forms an instrument for the study of political situation and of the distribution of political opinions, judgment and behavior at the approach of the 2007 elections. 5600 electors participated to the survey which is the first wave of four surveys. Along with the standard identifying information (age, sex, profession, etc.), it contains indicators of politicization, of relationship to politics, of the values and overall views of society held by the voters, of their perception of the issues and political questions and of their level of agreement with various political statements, we will denote these indicators "Feelings and beliefs variables". Naturally, this survey presents also the voting intentions for the former leader of extreme right in France: "Jean-Marie Le Pen". The representative nature of the sample was ensured via the quota method (sex cross-tabulated with age, occupation or former occupation of the head of household and level of educational qualifications) after stratification by administrative region and size of town.

2.1. The preselected data

In order to avoid very large models which could lead to erroneous estimation, we first choose relevant preselected variables:

- socioeconomic status: Level of education, Social position, Parents origins, Age, Sex, Residential area, Employment status, Family situation, Religion.
- Variables for feelings and beliefs about society and politics: Opinion on immigration, Problems in France, Opinion on unemployment, Opinion on homosexuality, Is France in decline, Favorite channel of Television, Opinion on globalization, Opinion on death sentence.

These sets of preselected variables contain the most common variables to explain the voting behavior (see Arzheimer [2]) and other specific to the available database which is very rich. However, the size of these sets is definitely too large to construct reasonable Bayesian network using these variables, we will now build an parsimonious explanatory model that employs standard logistic regression.

2.2. Model selection

Since our aim is to build a model by taking into account the plausible causality relationship between variables in order to avoid confounding variables and assesses precisely the influence of education level on Extreme-

Right voting we have too consider a lot of possible variables. However, now the number of preselected variables is too large and we have to begin by shrinking it by keeping only the most explicating ones. Hence, we select the most important socioeconomic variables then the most important belief variables. Since all the variables are categorical we can use loglinear models or, equivalently, logistic regression (see Agresti [1]). The binary logistic model is used to predict a binary response based on one or more predictor variables, in estimating the parameters of a qualitative response model. The probabilities describing the possible outcomes of a single trial are modeled, as a function of the explanatory variables, using a logistic function. The logit model with K explanatory factors $X_1, \cdots, X_K$ for the binary response Y may be written

$$\log\left(\frac{P\left(Y=1\,|X_1,\cdots X_K\,\right)}{1-P\left(Y=1\,|X_1,\cdots X_K\,\right)}\right)=\sum_{k=1}^{K}\beta_k^T X_k \tag{1}$$

where the I_k finite possible values of factor X_k are encoded on the simplex of $\mathbb{R}^{I_k}$ i.e. the vectors of $\mathbb{R}^{I_k}$ with one coordinate equal to 1 and 0 elsewhere. Another way of denoting such model is add an intercept (usually denoted α) and to add a constraint on the parameters $\beta_1, \cdots, \beta_K$ (for example the first component of β_1 will be set to zero) the writing of the model is then:

$$\log\left(\frac{P\left(Y=1\,|X_1,\cdots X_K\,\right)}{1-P\left(Y=1\,|X_1,\cdots X_K\,\right)}\right)=\alpha+\sum_{k=1}^{K}\beta_k^T X_k \tag{2}$$

but, for categorical explicating variables $X_1, \cdots, X_k$, this equation is equivalent of the previous writing. For $k \in \{1, \cdots, K\}$, $\beta_k^T X_k$ represents the effects of X_k through I_k parameters. This model assumes an absence of interaction in order to avoid the curse of dimensionality and a huge number of parameters. When a factor has no effect on the response variable, adjusting for the other factor, all its beta parameters are equal and has to be dropped. The criterion used to get the best set of variables will be the largely used BIC criterion:

$$-2\log(L) + K\log(n), \tag{3}$$

where L is the likelihood of the data, K is the number of free parameters and n is the number of observations. This is a type of criterion that Bozdogan [3] calls dimension consistent, since such criteria are often based on the assumption that a true model exists and this model is in the set of candidate models. Such criterion leads to a dimension estimate of this true model with probability 1 as sample size increases asymptotically. Other interpretations weaken this assumption to quasi-true models (see Cavanaugh and Neath [4]) and avoid the notion that BIC provides an estimate of the true models dimension but only the model nearest of the true one. Finally, with a Bayesian point of view, this criterion can be seen as an asymptotic approximation of the selection of the most probable models among all the models considered (see Raftery [9]).

2.2.1. Selection of socioeconomic variables

By comparing the BIC of all possible additive models we find the best set for socioeconomic variables:

- Level of education.
- Parents origins.
- Age.
- Residential area.

Although, some authors consider other variables, such selection seems to be coherent with principal texts on Extreme-Right vote (see for example Arzheimer [2]).

2.2.2. Selection of feelings and beliefs variables

Correspondingly we select the best additive logistic regression using only feelings and beliefs variables by comparing the BIC of all possible additive models. We obtain as best set of variables:

- Opinion on immigration.
- Opinion on globalization.
- Opinion on death sentence.

3. BAYESIAN NETWORK MODEL

A Bayesian network is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph. The edges represent conditional dependencies between the variables (nodes), and arrows represent causal relationships. We introduce this tool because this model allows to control confounding bias.

3.1. Controlling confounding bias

Whenever we undertake to evaluate the effect of one factor (X) on another (Y), the question arises as to whether we should adjust our measurements for possible variation in some other factors (Z) otherwise known as "confounders". Adjustment amounts to partitioning the population into groups that are homogeneous relative to Z, assessing the effect of X on Y in each homogeneous group, and then averaging the results. The illusive nature of such adjustment was recognized as early as 1899, when Karl Pearson discovered what is now called "Simpson's paradox" : Any statistical relationship between two variables may be reversed by including additional factors in the analysis. For example, we may find that students who smoke obtain higher grades than those who do not smoke but, adjusting for age, smoker obtain lower grades in every age group and, further adjusting for family income, smokers again obtain higher grades than nonsmoker in every income-age group, and so on. Despite a century of analysis, Simpson's reversal continues to "trap the unwary" (see David [6]), and the practical question that it poses: "whether an adjustment for a given covariate is appropriate", has resisted mathematical treatment. Since counterfactuals are not readily assertable from common scientific knowledge, the question has remained open: What criterion should one use to decide which variable are appropriate for adjustment? The next section a general solution of the adjustment problem using friendly langage of causal graphs.

3.2. The back-boor criterion

Assume we are given a causal graph G, together with non experimental data on a set V of observed variables which causal relationship are represented by G, and suppose we wish to estimate what effect the intervention for give the value x to a variable X would have on a response variable Y where X and Y are two variables of V. Pearl has shown (see [8]) that there exists a simple graphical test, named "back-door criterion" that can be applied directly to the causal diagram in order to test if a set of variables Z of V is sufficient for identifying the effect of X on Y. First we give a fundamental definition of blocked path in a graph G:

Definition 3.1 *A path is any unbroken route traced out the links between variables which may go either along or against the arrows (i.e. causal relationship) in a graph. We say that a path p is said blocked by a set Z of variables in G if and only if*

1. *p contains a chain $i \to m \to j$ or a fork $i \leftarrow m \to j$ such that the middle node (i.e. variable) m is in Z, or*
2. *p contains an inverted fork (or collider) $i \to m \leftarrow j$ such that the middle node m in not in Z and such that no descendant of m is in Z.*

These conditions mean that m block the flow of information along the path.

We can now give the definition of the back-door criterion:

Definition 3.2 *A set of variables Z satisfies the back-door criterion relative to an ordered pair of variables (X,Y) in a directed acyclic graph G if:*

- *No variable in Z is a descendant of X (i.e. is caused by X), and*
- *Z blocks every path between X and Y that contains an arrow into X.*

The name "back-door" echoes the second condition, which requires that only path with arrows pointing at X are blocked; these paths can be viewed as entering X through the back door. Finally we can compute the unbiased effect of X on Y by the theorem of back-door adjustment:

Theorem 3.1 *If a set of variables Z satisfies the back-door criterion relative to (X,Y), then the causal effect of X on Y is identifiable and is given by the formula:*

$$P(y\,|x\,) = \sum_{z} P(y\,|y, z\,)P(z) \tag{4}$$

The summation in (4) represents the standard formula obtained under adjustment for Z. In order to seek for the possible back doors in our data, we have first to built a Bayesian network model in the next section.

3.3. Estimated Bayesian structure

In order to construct the Bayesian network model, we use the R package "bnlearn" this package implements some algorithms for learning the structure of Bayesian networks. Constraint-based algorithms, also known as conditional independence learners, are all optimized derivatives of the Inductive Causation algorithm (Verma and Pearl, 1991). These algorithms use conditional independence tests to detect the Markov blankets of the variables, which in turn are used to compute the structure of the Bayesian network. We use our selected set of variables in order to build this network. Note that we set some reasonable constraints on the direction of possible causalities

$$\text{Socioeconomic Variables} \not\leftarrow \text{Feelings Beliefs Variables}$$

and

$$\left\{ \begin{array}{l} \text{Socioeconomic Variables} \\ \text{Feelings and Beliefs Variables} \end{array} \right. \not\leftarrow \text{Vote}$$

These constraint means that Belief variables cannot cause context variables and vote cannot causes Belief variables nor context variables.

Moreover since Belief variables are selected with the BIC criterion we assume that they are all useful to predict the Voting intention. Hence we get the constraint:

$$\text{Feelings and Beliefs Variables} \rightarrow \text{Vote}$$

Such constraint can be added in the call of bnlearn function using the blacklist, whitelist arguments:

- blacklisted arcs are never present in the graph.
- arcs whitelisted in one direction are always present in the graph.

We get finally the following model:

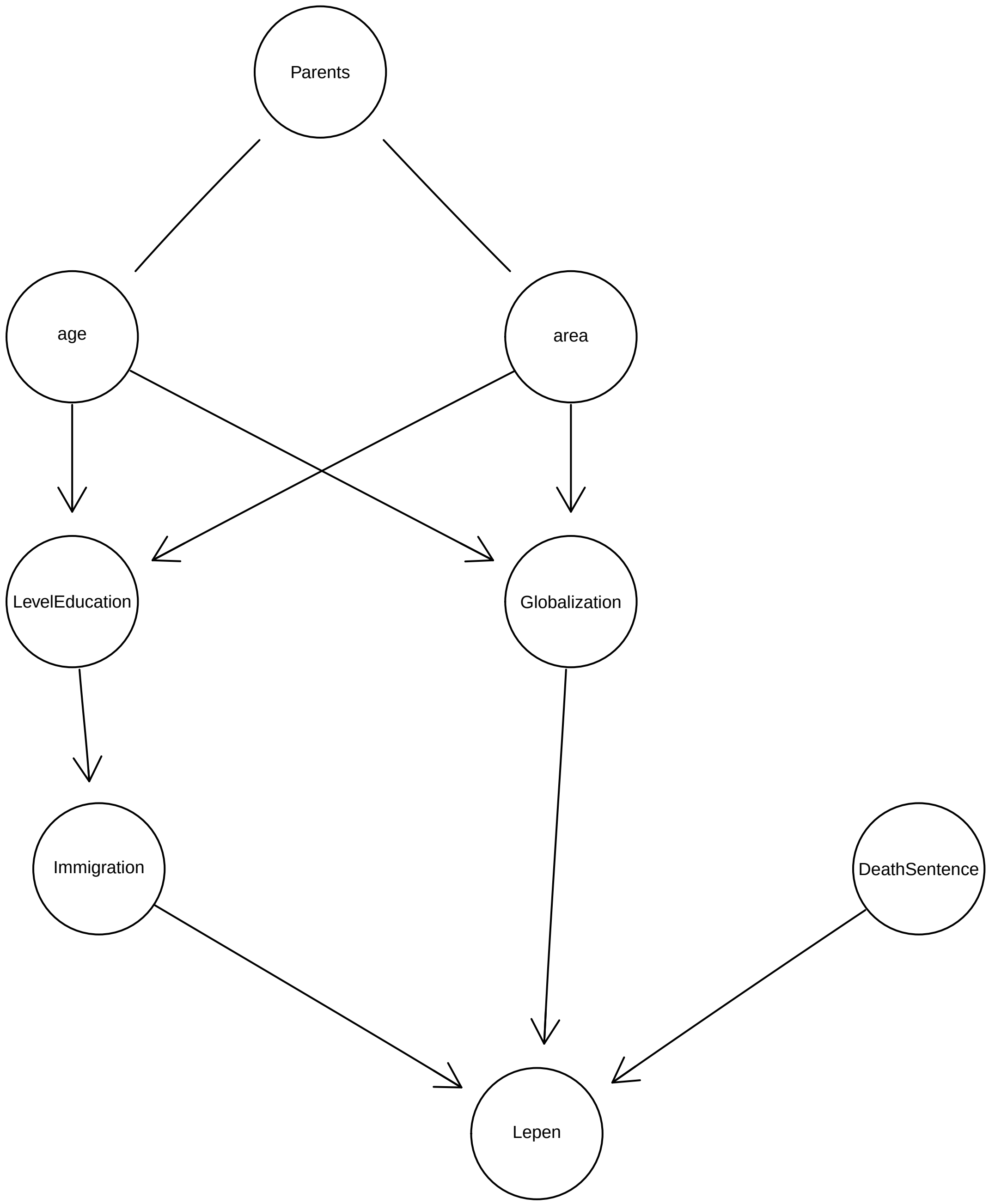

Figure 1: Bayesian network for voting intention

¿From this network we can estimate without bias of confounding variable the influence of education level on voting intention.

3.4. Estimation of influence of education

According to Pearl [8] we see that the variable G : “Opinion on globalization” is a back door, hence the influence of the variable E: “Level of education” on the voting intention for “Le Pen” is identifiable and can be computed with a control of confounding bias as:

$$P\left(Lepen = 1\left|E = e\right.\right) = \sum_{g \in I_G} P\left(Lepen = 1\left|E = e, G = g\right.\right) \times P(G = g)$$

where I_G is the set of possible opinion on globalization (an opportunity, a threat or without opinion) and the levels of education are:

- $e = 1$: Without diploma.
- $e = 2$: Before high school
- $e = 3$: High school
- $e = 4$: Under graduate
- $e = 5$: Graduate

The probability of vote is estimated by:

$$\begin{array}{l} \hat{P}\left(Lepen = 1\left|E = e\right.\right) = \sum_{g \in I_g} \hat{P}\left(Lepen = 1\left|E = e, G = g\right.\right) \times \hat{P}(G = g) \\ = \sum_{g \in I_g} \frac{\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{1,e,g\}}(Lepen_t, E_t, G_t)}{\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{e,g\}}(E_t, G_t)} \times \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{g\}}\left(G_t\right), \end{array} \tag{5}$$

where $\mathbf{1}$ is the indicator function. This estimate is a function of

$$\hat{P}(e) = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{1,e,1\}}\left(Lepen_t, E_t, G_t\right) \\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{1,e,2\}}\left(Lepen_t, E_t, G_t\right) \\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{1,e,3\}}\left(Lepen_t, E_t, G_t\right) \\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{e,1\}}\left(E_t, G_t\right) \\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{e,2\}}\left(E_t, G_t\right) \\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{e,3\}}\left(E_t, G_t\right) \\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{1\}}\left(G_t\right) \\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{2\}}\left(G_t\right) \end{pmatrix}$$

where

$$\hat{P}(Lepen = 1|E = e) = h\left(\hat{P}(e)\right).$$

With

$$h\left(\begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ g_1 \\ g_2 \end{pmatrix}\right) = \frac{z_1}{z_4}\, g_1 + \frac{z_2}{z_5}\, g_2 + \frac{z_3}{z_6}\, (1 - (g_1 + g_2)). \tag{6}$$

Since $\hat{P}(e)$ is an unbiased estimate of

$$P(e) = \begin{pmatrix} P\left(Lepen = 1, E = e, G = 1\right) \\ P\left(Lepen = 1, E = e, G = 2\right) \\ P\left(E = e, G = 3\right) \\ P\left(E = e, G = 1\right) \\ P\left(E = e, G = 2\right) \\ P\left(E = e, G = 3\right) \\ P\left(G = 1\right) \\ P\left(G = 2\right) \end{pmatrix}$$

by the delta method, we get:

$$\sqrt{n}\left(h\left(\hat{P}(e)\right) - h\left(P(e)\right)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\ J_h(P(e))\,\Sigma J_h^T(P(e))\right) \tag{7}$$

where $J_h(P(e))$ is the Jacobian matrix of h in $P(e)$ and Σ is the variance-covariance matrix of the variable

$$V = \begin{pmatrix} \mathbf{1}_{\{1,e,1\}}(Lepen, E, G) \\ \mathbf{1}_{\{1,e,2\}}(Lepen, E, G) \\ \mathbf{1}_{\{1,e,3\}}(Lepen, E, G) \\ \mathbf{1}_{\{e,1\}}(E, G) \\ \mathbf{1}_{\{e,2\}}(E, G) \\ \mathbf{1}_{\{e,3\}}(E, G) \\ \mathbf{1}_{\{1\}}(G) \\ \mathbf{1}_{\{2\}}(G) \end{pmatrix}$$

We can get an estimation of this quantity by plugging the estimation of $P(e)$ and Σ into this formula:

$$J_h(P(e))\,\Sigma J_h^T(P(e)) = J_h(\hat{P}(e))\hat{\Sigma} J_h^T(\hat{P}(e)) \tag{8}$$

The details of computation is left in the appendix section.

3.5. Numerical results

Using the previous formula, we get the following numbers:

Results with Back-Door			
E	$\hat{P}\left(Lepen = 1/E = e\right)$	$\left(J_h\left(\hat{P}\left(e\right)\right)\hat{\Sigma}J_h^T\left(\hat{P}\left(e\right)\right)\right)/n$	$CI\left(P\left(Lepen = 1/E = e\right)\right)$
1	0.1366141	7.27293e-05	[0.1198990 , 0.1533293]
2	0.1207164	4.961633e-05	[0.1069103 , 0.1345224]
3	0.06180464	6.731698e-05	[0.04572344 , 0.07788584]
4	0.05563439	0.0001125881	[0.03483732 , 0.07643147]
5	0.02568848	5.908753e-05	[0.01062227 , 0.04075469]

Hence, the probability of vote for Le Pen can be parted into three sets:

1. High probability (around 13%) for low level of education.
2. Medium probability (around 6%) for medium level of education.
3. Low probability (less than 3%) for high level of education.

By comparison we give the biased results computed without taking into account the back-door adjustment, note that this estimator is biased for estimate the causal effect of education on the voting intention but its computation and the computation of it variance with the δ-method is simple and straightforward:

Hence for $e\epsilon\{1, 2, 3, 4, 5\}$, this estimator is:

$$\hat{P}_W(Lepen = 1\,|E = e) = \frac{\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{1,e}(Lepen_t, E_t)}{\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_e(E_t)} \tag{9}$$

This estimate is a function of

$$\hat{P}_W(e) = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{1,e\}}\left(Lepen_t, E_t\right) \\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{e\}}\left(E_t\right) \end{pmatrix} \tag{10}$$

where

$$\hat{P}_W(Lepen = 1|E = e) = f\left(\hat{P}(e)\right)$$

with

$$f\left(\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right) = \frac{z_1}{z_2}. \tag{11}$$

$\hat{P}(e)$ is an empirical mean that estimates

$$P_W(e) = \begin{pmatrix} P\left(Lepen = 1, E = e\right) \\ P\left(E = e\right) \end{pmatrix}.$$

By the delta method, we get:

$$\sqrt{n}\left(h\left(\hat{P}_W(e)\right) - h\left(P_W(e)\right)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\ J_h(P_W(e))\,\Sigma_W J_h^T(P_W(e))\right) \tag{12}$$

$J_f(P_W(e))$ is the Jacobian matrix of f in $P_W(e)$ and Σ_W is the variance-covariance matrix of the variable

$$V_W = \begin{pmatrix} \mathbf{1}_{\{1,e\}}(Lepen, E) \\ \mathbf{1}_{\{e\}}(E) \end{pmatrix}$$

We can get an estimation of this quantity by plugging the estimation of $P(e)$ and Σ into this formula:

$$J_f(P_W(e))\,\Sigma J_f^T(P_W(e)) = J_f(\hat{P}_W(e))\hat{\Sigma}_W J_f^T(\hat{P}_W(e)) \tag{13}$$

Results without the Back-Door adjustment			
E	$\hat{P}_W\left(Lepen = 1/E = e\right)$	$\left(J_f\left(\hat{P}_W\left(e\right)\right)\hat{\Sigma}_W J_f^T\left(\hat{P}_W\left(e\right)\right)\right)/n$	$CI\left(P_W\left(Lepen = 1/E = e\right)\right)$
1	0.1435006	0.000105594	[0.1233599 , 0.1636414]
2	0.1244019	6.692715e-05	[0.1083673 , 0.1404365]
3	0.06050228	7.324522e-05	[0.04372793 , 0.07727663]
4	0.05182342	0.000104624	[0.03177539 , 0.07187144]
5	0.02184874	3.752287e-05	[0.009842581 , 0.033854898]

We can see, that the estimation of the influence of the education is slightly overestimated in this computation. The less educated classes seem to vote more for the Extreme Right and the most educated class to vote less for Extreme Right even if the difference is small with the previous results.

4. CONCLUSION

A Bayesian network has been built to assess the influence of the level of education on the voting intention for the “Le Pen” candidate. This model aims to take the possible confusion variables into account. It appears that the variable “Opinion on globalization” is a fundamental one. Indeed, according to the theory of Bayesian network, this variable satisfies the back door criterion, hence the causal effect of level of education on the vote intention is identifiable and can be computed with a control of possible confounding bias. Here, no “Simpson paradox” appears but, even if the level of education is a fundamental variable influencing the voting intention for Extreme Right, it effect seems to be a little bit overestimated if confounding variables are overlooked. In a future work, it would be interesting to confirm such findings on more recent data involving the new leader of the Extreme right in France: “Marine Le Pen” the daughter of “Jean-Marie Le Pen”.

RECEIVED: JULY, 2015.
REVISED: DECEMBER, 2015.

REFERENCES

[1] AGRESTI, A. [1990]: **Categorical data analysis** John Wiley & Sons, New York.

[2] ARZHEIMER, K. [2009]: Contextual factors and the extreme right vote in Western Europe, 1980-2002 **American Journal of Political Science**, 53, 2,:259–275.

[3] BOZDOGAN, H. [1987]: Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions **Psychometrika**, 52:345–370.

[4] CAVANAUGH, J. and NEATH, A. [1999]: Generalizing the derivation of the schwarz information criterion **Communication in statistic Theory and methods**, 28:49–66.

[5] (CEVIPOF), C. D. E. D. L. V. P. F. [2007]: Barométre politique francais 2006-2007 **Econometrica**.

[6] DAWID, A. [1979]: Conditional independence in statistical theory **Journal of the Royal statistical Society, Series B**, 41, 1,:1–31.

[7] MÜLLER, J. [1999]: From national identity to national interest: The rise (and fall) of Germany's new right **German Politics**, 8, 3,.

[8] PEARL, J. [2000]: **Causality: Models, reasoning and inference** Cambridge University Press, Cambridge, MA, USA, second edition.

[9] RAFTERY, A. E. [1995]: Bayesian model selection in social research **Sociological methodology**, 25:111–163.

[10] SCHUMAN, H., S. C. B. L. and KRYSAN, M. [1997]: **Racial attitudes in America**, volume 28.

[11] SCUTARI, M. [2010]: Learning bayesian networks with the BNlearn R package. **Journal of statistical software,**, 35.

[12] VOGT, P. [1997]: **Tolerance and education. Learning to live with diversity and difference.**

5. APPENDIX

Let us write

$$\begin{cases} P_{1,e,1} = P\left(Lepen = 1, E = e, G = 1\right) \\ P_{1,e,2} = P\left(Lepen = 1, E = e, G = 2\right) \\ P_{1,e,3} = P\left(E = e, G = 3\right) \\ P_{e,1} = P\left(E = e, G = 1\right) \\ P_{e,2} = P\left(E = e, G = 2\right) \\ P_{e,3} = P\left(E = e, G = 3\right) \\ P_{1} = P\left(G = 1\right) \\ P_{2} = P\left(G = 2\right) \end{cases}$$

We get for all $i \in I_g \text{and} j \in I_g$:

$$
\begin{aligned}
&Cov\left(\mathbf{1}_{\{1,e,i\}}(Lepen,E,G),\mathbf{1}_{\{1,e,j\}}(Lepen,E,G)\right)=\begin{cases}-P_{1,e,i}P_{1,e,j} & i\neq j\\ P_{1,e,i}(1-P_{1,e,i}) & i=j\end{cases}\\
&Cov\left(\mathbf{1}_{\{1,e,i\}}(Lepen,E,G),\mathbf{1}_{\{e,j\}}(E,G)\right)=\begin{cases}-P_{1,e,i}P_{e,j} & i\neq j\\ P_{1,e,i}(1-P_{e,i}) & i=j\end{cases}\\
&Cov\left(\mathbf{1}_{\{1,e,i\}}(Lepen,E,G),\mathbf{1}_{\{j\}}(G)\right)=\begin{cases}-P_{1,e,i}P_{j} & i\neq j\\ P_{1,e,i}(1-P_{i}) & i=j\end{cases}\\
&Cov\left(\mathbf{1}_{\{e,i\}}(E,G),\mathbf{1}_{\{e,j\}}(E,G)\right)=\begin{cases}-P_{e,i}P_{e,j} & i\neq j\\ P_{e,i}(1-P_{e,i}) & i=j\end{cases}\\
&Cov\left(\mathbf{1}_{\{e,i\}}(E,G),\mathbf{1}_{\{j\}}(G)\right)=\begin{cases}-P_{e,i}P_{j} & i\neq j\\ P_{e,i}(1-P_{i}) & i=j\end{cases}\\
&Cov\left(\mathbf{1}_{\{i\}}(G),\mathbf{1}_{\{j\}}(G)\right)=\begin{cases}-P_{i}P_{j} & i\neq j\\ P_{i}(1-P_{i}) & i=j\end{cases}
\end{aligned}
\tag{14}
$$

Hence

$$
\Sigma=(\Sigma_1,\Sigma_2,\Sigma_3,\Sigma_4,\Sigma_5,\Sigma_6,\Sigma_7,\Sigma_8)=\begin{pmatrix}\Sigma_1^T\\ \Sigma_2^T\\ \Sigma_3^T\\ \Sigma_4^T\\ \Sigma_5^T\\ \Sigma_6^T\\ \Sigma_7^T\\ \Sigma_8^T\end{pmatrix}
$$

with

$$
\begin{aligned}
&\Sigma_1^T=(P_{1,e,1}(1-P_{1,e,1}),-P_{1,e,1}P_{1,e,2},-P_{1,e,1}P_{1,e,3},P_{1,e,1}(1-P_{e,1}),-P_{1,e,1}P_{e,2},\\
&-P_{1,e,1}P_{e,3},P_{1,e,1}(1-P_1),-P_{1,e,1}P_2)\\
&\Sigma_2^T=(-P_{1,e,2}P_{1,e,1},P_{1,e,2}(1-P_{1,e,2}),-P_{1,e,2}P_{1,e,3},-P_{1,e,2}P_{e,1},P_{1,e,2}(1-P_{e,2}),\\
&-P_{1,e,2}P_{e,3},-P_{1,e,2}P_1,P_{1,e,2}(1-P_2))\\
&\Sigma_3^T=(-P_{1,e,3}P_{1,e,1},-P_{1,e,3}P_{1,e,2},P_{1,e,3}(1-P_{1,e,3}),-P_{1,e,3}P_{e,1},-P_{1,e,3}P_{e,2},\\
&P_{1,e,3}(1-P_{e,3}),-P_{1,e,3}P_1,-P_{1,e,3}P_2)\\
&\Sigma_4^T=(P_{1,e,1}(1-P_{e,1}),-P_{e,1}P_{1,e,2},-P_{e,1}P_{1,e,3},P_{e,1}(1-P_{e,1}),-P_{e,1}P_{e,2},\\
&-P_{e,1}P_{e,3},P_1(1-P_{e,1}),-P_{e,1}P_2)\\
&\Sigma_5^T=(-P_{e,2}P_{1,e,1},P_{1,e,2}(1-P_{e,2}),-P_{e,2}P_{1,e,3},-P_{e,2}P_{e,1},P_{e,2}(1-P_{e,2}),\\
&-P_{e,2}P_{e,3},-P_{e,2}P_1,P_{e,2}(1-P_2))\\
&\Sigma_6^T=(-P_{e,3}P_{1,e,1},-P_{e,3}P_{1,e,2},P_{1,e,3}(1-P_{e,3}),-P_{e,3}P_{e,1},-P_{e,3}P_{e,2},\\
&P_{e,3}(1-P_{e,3}),-P_{e,3}P_1,-P_{e,3}P_2)\\
&\Sigma_7^T=(P_{1,e,1}(1-P_1),-P_1P_{1,e,2},-P_1P_{1,e,3},P_{e,1}(1-P_1),-P_1P_{e,2},-P_1P_{e,3},\\
&P_1(1-P_1),-P_1P_2)\\
&\Sigma_8^T=(-P_2P_{1,e,1},P_2(1-P_{1,e,2}),-P_2P_{1,e,3},-P_2P_{e,1},P_2(1-P_{e,2}),-P_2P_{e,3},\\
&-P_2P_1,P_2(1-P_2))\,.
\end{aligned}
\tag{15}
$$

Then, we get the empirical estimation $\hat{\Sigma}$ by replacing all P by their empirical estimations $\hat{P}$