

# Influence d'une chaîne de TV conservatrice sur le vote aux élections présidentielles américaines

Jérémy L'Hour  
jeremy.l.hour@ensae.fr

June 6, 2018

Ce projet tente de quantifier l'effet de la diffusion d'une chaîne de télévision conservatrice entre 1996 et 2000 sur le vote en faveur d'un parti conservateur aux élections présidentielles américaines correspondantes. MAGA News est une chaîne d'information en continu créée en septembre 1996. Les spécificités du marché américain de la télévision câblée rendent le déploiement d'une nouvelle chaîne de TV nécessairement long et coûteux : étant donnés les coûts fixes, chaque ville américaine constitue un monopole local opéré par une des quelques grandes compagnies de télévision câblée, qui font face à des contraintes sur le nombre de chaînes qu'elles peuvent diffuser. MAGA News a donc dû négocier sa diffusion auprès des compagnies de câble, souvent au détriment d'autres chaînes. Elle est considérée comme une chaîne à droite sur le spectre politique américain, dans l'absolu et relativement aux chaînes de télévision concurrentes.

**Données** La base de données `MAGANews.dta` regroupe des informations concernant la diffusion des chaînes de télé, les caractéristiques sociodémographiques et le vote, mesurées pour 9,265 villes américaines sur plusieurs années. Elle est disponible à l'adresse [www.github.com/jlhourENSAE/MAGA-Econometrics](https://www.github.com/jlhourENSAE/MAGA-Econometrics). Dans la suite de l'énoncé, on appellera "groupe traité" l'ensemble des villes ayant accès à MAGA News en 2000 et "groupe contrôle" l'ensemble des autres villes.

**Consignes importantes** Pour toute réponse demandant une estimation, vous devez fournir un écart-type correspondant, dont le mode de calcul sera justifié. Une grande attention sera

portée à la façon d’exposer vos résultats. Inspirez-vous notamment de la façon dont les articles de recherche reportent les résultats de régression.

**Ressources utiles** Concernant les aspects statistiques de ce projet, “Large Sample Estimation and Hypothesis Testing” de Newey et McFadden (*Handbook of Econometrics*, 1994); à propos du score de propension, la partie 3 du livre *Causal Inference for Statistics, Social and Biomedical Sciences* de Imbens et Rubin (2015).

## 1. Statistiques Descriptives et Double-Différences

1. Quelle est la proportion des villes dans lesquelles MAGA News est diffusée en 1998, 2000 et 2003 ? En quoi cela offre-t-il la possibilité de mesurer l’impact de MAGA News sur le vote ?
2. La variable `reppresfv2p1996` (resp. `reppresfv2p2000`) mesure la part du vote conservateur aux élections présidentielles de 1996 (resp. 2000). Calculer l’estimateur des différences-de-différences de l’effet de MAGA News.
3. L’estimateur précédent est-il crédible ? Effectuer un ou plusieurs tests pour justifier votre réponse.
4. Produire des statistiques descriptives sur les caractéristiques des villes permettant d’éclairer la réponse à la question précédente.

## 2. Sélection et Score de Propension

Le but de cette section est d’étudier les facteurs déterminants de la présence de MAGA News dans une ville (*i.e.* la sélection dans le traitement). Pour cette partie, vous devrez considérer des variables de trois natures différentes : (1) économiques, liées au marché local de la télévision câblée, (2) liées au vote et à la couleur politique des citoyens d’une ville, (3) sociodémographiques. Plus précisément, le but de cette partie est de montrer que l’implémentation de MAGA News dans une ville n’est pas liée à deux facteurs que sont

(i) la part du vote conservateur à l'élection de 1996 (`reppresfv2p1996`) et (ii) le taux de participation à l'élection de 1996 (`totpreslvpop1996`).

1. En quoi serait-il problématique que les deux variables citées précédemment expliquent la présence de MAGA News en 2000 ?
2. Proposer quatre spécifications différentes, avec un nombre croissant de variables, d'un modèle Logit où la variable expliquée est la présence de MAGA News dans une ville en 2000. La première spécification ne fera intervenir que la part du vote conservateur à l'élection de 1996 et le taux de participation à l'élection de 1996. Vous justifierez les variables de contrôle que vous ajoutez, mais pour ces variables uniquement, on ne demande pas de reporter les résultats de l'estimation. On prêtera attention, via des tests statistiques pertinents, à la qualité des modèles proposés.
3. D'après votre réponse à la question précédente, la part du vote conservateur à l'élection de 1996 et le taux de participation à l'élection de 1996 expliquent-ils l'implémentation de MAGA News ? Faire les tests correspondant.
4. Soit  $D_i$  la variable aléatoire qui vaut un si MAGA News est disponible dans la ville  $i$  en 2000 et zéro sinon, et  $X_i$  un vecteur aléatoire de dimension  $p$  mesurant des variables explicatives.

(a) On définit la variable aléatoire:

$$W_i := \frac{\mathbb{E}(1 - D_i)}{\mathbb{E}(D_i)} \exp(X_i' \beta_0).$$

Montrez que si le score de propension est donné par un modèle Logit (*i.e.*  $\mathbb{P}[D = 1|X] = [1 + \exp(-X' \beta_0)]^{-1}$ ) alors:

$$\mathbb{E}[X_i | D_i = 1] = \mathbb{E}[W_i X_i | D_i = 0]. \quad (1)$$

(b) Interpréter cette équation.

5. (a) Montrer que l'équation 1 peut s'écrire comme une condition de moment du type  $\mathbb{E}[g(D_i, X_i, \beta_0)] = 0$  où vous préciserez la fonction  $g$ .

- (b) Proposer un estimateur GMM (Méthode des Moments Généralisée) de  $\beta_0$ . Le calculer pour les quatre spécifications choisies à la question 2.

### 3. Estimation d'Impact par Régression Linéaire

Le but de cette section est d'estimer l'impact causal de la présence de MAGA News sur le vote conservateur au moyen de régressions linéaires, afin d'obtenir un estimateur meilleur que celui des différences-de-différences obtenu dans la partie 1.

*La variable `totpresvotes1996` donne le nombre de votes exprimés pour l'élection présidentielle de 1996. Dans cette section, on ponderera les résultats d'estimation par cette variable, de façon à interpréter les résultats pour l'électeur moyen, plutôt que pour la ville moyenne.*

1. En prenant en compte les résultats statistiques de la partie 2, proposez quatre spécifications de la régression linéaire de la différence entre la part du vote conservateur à l'élection de 2000 et celle de 1996 (mesurée par `reppresfv2p00m96`) sur la présence de MAGA News en 2000 (mesurée par `maganews2000`). Justifiez :
  - (a) la sélection des variables de contrôle,
  - (b) la présence ou non d'une variable mesurant la différence entre la part du vote conservateur entre deux élections avant 1996,
  - (c) l'utilisation d'effets fixes,
  - (d) le calcul de l'écart-type.

*Le report des résultats d'estimation pour les coefficients associés aux variables de contrôle n'est pas demandé.*

2. On note  $Y_i(1)$  (resp.  $Y_i(0)$ ) la variable aléatoire `reppresfv2p00m96` qui mesure la différence entre la part du vote conservateur à l'élection de 2000 et celle de 1996 quand MAGA News est disponible dans la ville  $i$  (resp. quand MAGA News n'est pas disponible dans la ville  $i$ ). On note la variable observée  $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$ . On définit le paramètre  $\theta_0$  :

$$\theta_0 = \mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0]. \quad (2)$$

On suppose que  $Y(0) \perp\!\!\!\perp D|X$  (*Hypothèse d'Indépendance Conditionnelle*, ou CIA).

- (a) Montrer que  $\theta_0 = \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]$ .
  - (b) Proposer un estimateur de  $\theta_0$  basé sur l'équation (2), que l'on notera  $\hat{\theta}$ .
  - (c) Montrer que  $\hat{\theta}$  est asymptotiquement normal. Donner son écart-type asymptotique.
  - (d) A partir de l'estimateur de  $\beta_0$  calculé à la Q5 de la partie 2, calculer  $\hat{\theta}$  ainsi que son écart-type. Comparer aux résultats obtenus par régression.
3. La disponibilité de MAGA News sur le réseau de télévision local impacte-t-il le vote conservateur ? Dans quelle mesure ? Proposez deux explications à ce phénomène (ou à son absence).

#### 4. Hétérogénéité de l'effet et Test Placebo

1. On souhaite étudier l'existence d'effets hétérogènes de MAGA News. Pour une spécification choisie dans la partie 3, on fera interagir la présence de MAGA News avec :
- (a) Le nombre de chaîne disponibles sur le réseau de la ville en 2000,
  - (b) La part de la population résidant en zone urbaine,
  - (c) Le niveau d'éducation de la population.
- On réalisera une ou plusieurs régressions pour chacun des trois items. La spécification (variable en continu, par tranches, indicatrices, sous-échantillons etc.) est laissée libre.*
2. On souhaite conduire un test placebo. Pour cela, on va regarder l'impact de la présence de MAGA News en 2000 sur l'évolution de la part du vote conservateur entre 1992 et 1996, ainsi qu'entre 1988 et 1992. Quel effet devrait-on observer ? Mettre en œuvre ce test pour une spécification choisie. Cela vous donne-t-il plus de confiance dans les résultats obtenus ?

## 5. Elements de Correction

Pour la partie empirique, voir l'article “*The Fox News Effect: Media Bias and Voting*” de Stefano DellaVigna et Ethan Kaplan (QJE, 2007). Certains labels ont été modifiées pour ce projet, il s’agissait non pas de MAGA News qui est un nom fictif, mais bien de Fox News.

**Double Différences.** L’estimateur de double-différences s’utilise dans un cadre où il existe deux groupes ( $D_i = 0, 1$  selon que la ville appartienne ou non au groupe des villes ayant accès à Fox News en 2000) et deux dates ( $t = 1996, 2000$ ). Le groupe contrôle regroupant les villes telles que  $D_i = 0$  n’est jamais traité. Tandis que le groupe traité  $D_i = 1$  est traité en 2000 mais pas en 1996. Soit  $Y_{i,t}$ , la part du vote républicain dans la ville  $i$  à la date  $t$ , le modèle classique est:

$$Y_{i,t} = \delta_t + D_i\tau_t + \varepsilon_{i,t}, \quad (3)$$

où  $\delta_t$  désigne l’effet fixe temporel pour le groupe contrôle, tandis que  $\delta_t + \tau_t$  désigne l’effet fixe temporel pour le groupe traité. Le modèle de différence de différences indique que l’effet du traitement est donné par  $\tau_{2000} - \tau_{1996}$ . Une régression linéaire de  $Y_{i,t}$  sur une indicatrice de l’année 2000 (coefficient associé  $\delta_{2000}$ ),  $D_i$  (coefficient associé  $\tau_{1996}$ ) et l’interaction de ces deux variables (coefficient associé  $\tau_{2000} - \tau_{1996}$ ) permet d’estimer le modèle. On inclut une constante dans la régression, qui permet d’estimer le coefficient  $\delta_{1996}$ . Il est à noter que les écarts-types seront potentiellement à corriger pour prendre en compte la corrélation intra-ville (*i.e.* entre deux observations d’une même ville aux deux dates). On pouvait également considérer le modèle en différence première:

$$\Delta Y_{i,2000} = Y_{i,2000} - Y_{i,1996} = \delta_{2000} - \delta_{1996} + D_i(\tau_{2000} - \tau_{1996}) + \Delta \varepsilon_{i,2000}, \quad (4)$$

directement estimable par une régression de  $\Delta Y_{i,2000}$  sur  $D_i$ , où les écarts-types seront correctement estimés a priori car à garder l’hypothèse d’indépendance des résidus entre les villes, il n’y a pas besoin de cluster. A noter que ces deux régressions vont donner rigoureusement la même estimation du coefficient. Pour approfondir, voir “*Mostly Harmless Econometrics*”, Chapitre 5.2.

Un erreur classique a été de confondre  $D_i$  (indicatrice d'appartenance à un groupe) et “avoir accès à Fox News” (indicatrice de traitement, que l'on note  $T_{i,t}$ , qui change de valeur selon  $i$  ET  $t$ ). C'est bien  $D_i$  qu'utilise l'estimateur de différence-de-différences. Utiliser une indicatrice de traitement à la place de  $D_i$  ne permettait pas d'aboutir à un estimateur à cause de la multicollinéarité des variables, car  $T_{i,t} = D_i 1 \{\text{on est à la période } t\} = T_{i,t} 1 \{\text{on est à la période } t\}$ .

L'hypothèse centrale pour conclure à la nature causale de l'estimateur de différence-de-différences est l'hypothèse de tendance commune dont on pouvait apprécier la pertinence au moyen de régressions utilisant les élections de 1988, 1992 et 1996.

**Partie 2: Q4-Q5** Soit  $D_i$  la variable aléatoire qui vaut un si Fox News est disponible dans la ville  $i$  en 2000 et zéro sinon, et  $X_i$  un vecteur aléatoire de dimension  $p$  mesurant des variables explicatives. On définit la variable aléatoire:

$$W_i := \frac{\mathbb{E}(1 - D_i)}{\mathbb{E}(D_i)} \exp(X_i' \beta_0).$$

Le score de propension est défini par un Logit (*i.e.*  $\mathbb{P}[D = 1|X] = [1 + \exp(-X' \beta_0)]^{-1}$ ).

Nous avons donc:

$$\begin{aligned} \mathbb{E}[W_i X_i | D_i = 0] &= \frac{\mathbb{E}(1 - D_i)}{\mathbb{E}(D_i)} \mathbb{E}[\exp(X_i' \beta_0) X_i | D_i = 0] \\ &= \frac{1}{\mathbb{E}(D_i)} \mathbb{E}[(1 - D_i) \exp(X_i' \beta_0) X_i] \\ &= \frac{1}{\mathbb{E}(D_i)} \mathbb{E}[\mathbb{E}(1 - D_i | X_i) \exp(X_i' \beta_0) X_i] \\ &= \frac{1}{\mathbb{E}(D_i)} \mathbb{E}\left[\frac{1}{1 + \exp(X_i' \beta_0)} \exp(X_i' \beta_0) X_i\right] \\ &= \frac{1}{\mathbb{E}(D_i)} \mathbb{E}[\mathbb{E}(D_i | X_i) X_i] \\ &= \mathbb{E}[X_i | D_i = 1]. \end{aligned}$$

L'équation 1 s'interprète comme une condition d'équilibre: le poids  $W_i$  permet de repondérer le groupe de contrôle de sorte qu'il se retrouve avec les mêmes caractéristiques moyennes que le groupe traité. Normalement, cela doit permettre de répondre au déséquilibre observé à la fin de la partie 1.

On prend  $g(D_i, X_i, \beta_0) = (D_i - (1 - D_i) \exp(X_i' \beta_0)) X_i$ . En effet d'après la condition d'équilibre précédente:

$$\begin{aligned} \mathbb{E}[D_i X_i] &= \mathbb{E}(D_i) \mathbb{E}[X_i | D_i = 1] \\ &= \mathbb{E}(D_i) \mathbb{E}[W_i X_i | D_i = 0] \\ &= \frac{\mathbb{E}(D_i)}{\mathbb{E}(1 - D_i)} \mathbb{E}[(1 - D_i) W_i X_i] \\ &= \mathbb{E}[(1 - D_i) \exp(X_i' \beta_0) X_i], \end{aligned}$$

ce qui prouve que  $\mathbb{E}[g(D_i, X_i, \beta_0)] = 0$ .

On cherche à estimer  $\beta_0$  qui est un paramètre de dimension  $p$ .  $\mathbb{E}[g(D_i, X_i, \beta_0)] = 0$  représente  $p$  conditions de moments, que l'on peut utiliser pour estimer ce paramètre. On a alors des GMM juste identifiés - la matrice de pondération des moments n'importe pas. On peut également combiner ces  $p$  moments à  $p$  autres moments, provenant du score du modèle logit. En effet, on peut estimer  $\beta_0$  par maximum de vraisemblance:

$$\max_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n D_i X_i' \beta - \log(1 + \exp(X_i' \beta)),$$

dont la condition du premier ordre est:

$$\frac{1}{n} \sum_{i=1}^n D_i X_i - \frac{\exp(X_i' \beta)}{1 + \exp(X_i' \beta)} X_i = 0.$$

Cela correspond aux  $p$  conditions de moments  $\mathbb{E}[s(D_i, X_i, \beta_0)] = 0$  pour  $s(D_i, X_i, \beta_0) = \left(D_i - \frac{\exp(X_i' \beta_0)}{1 + \exp(X_i' \beta_0)}\right) X_i$ . Dans ce cas, on a des GMM sur-identifiés, car on a  $2p$  conditions de moments pour  $p$  paramètres. Le calcul des écart-types peut se faire via les formules données dans le cours, ou bien est automatiquement dans la plupart des logiciels habituels (Stata, R).

**Partie 3: Q2** Il s'agit de montrer que  $\mathbb{E}[W_i Y_i | D_i = 0] = \mathbb{E}[Y_i(0) | D_i = 1]$ , i.e. que la moyenne pondérée de l'outcome pour le groupe de contrôle fournit un bon contrefactuel. Notons que l'hypothèse d'indépendance conditionnelle implique que  $\mathbb{E}((1 - D_i) Y_i(0) | X_i) = \mathbb{E}(1 - D_i | X_i) \mathbb{E}(Y_i(0) | X_i)$ .

$$\mathbb{E}[W_i Y_i | D_i = 0] = \mathbb{E}[W_i Y_i(0) | D_i = 0]$$



$$\begin{aligned}
&= \frac{1}{\mathbb{E}(D_i)} \mathbb{E} [(1 - D_i) \exp(X_i' \beta_0) Y_i(0)] \\
&= \frac{1}{\mathbb{E}(D_i)} \mathbb{E} [\exp(X_i' \beta_0) \mathbb{E} ((1 - D_i) Y_i(0) | X_i)] \\
&= \frac{1}{\mathbb{E}(D_i)} \mathbb{E} \left[ \frac{\exp(X_i' \beta_0)}{1 + \exp(X_i' \beta_0)} \mathbb{E} (Y_i(0) | X_i) \right] \\
&= \frac{1}{\mathbb{E}(D_i)} \mathbb{E} [\mathbb{E} (D_i | X_i) \mathbb{E} (Y_i(0) | X_i)] \\
&= \mathbb{E} [Y_i(0) | D_i = 1].
\end{aligned}$$

On a par ailleurs, facilement que  $\mathbb{E} [Y_i | D_i = 1] = \mathbb{E} [Y_i(1) | D_i = 1]$ . On notera au passage qu'on ne pouvait pas appliquer directement le résultat sur l'équation d'équilibre car  $Y_i$  n'est pas incluse dans le vecteur  $X_i$ : l'équilibre ne tient que pour les  $X_i$  entrant dans le score de propension et non pour l'outcome  $Y_i(0)$  *directement*. Enfin, on notera que  $Y(0)$  n'est pas linéaire en  $X$ . Une autre démonstration est tout de même possible. Notons que l'on peut toujours écrire:  $Y_i(0) = \mathbb{E} [Y_i(0) | X_i, D_i] + \eta_i$  avec une variable aléatoire  $\eta_i$  telle que  $\eta_i \perp\!\!\!\perp (X_i, D_i)$ . La CIA implique que  $\mathbb{E} [Y_i(0) | X_i, D_i] = \mathbb{E} [Y_i(0) | X_i]$ . Dès lors:

$$\begin{aligned}
\mathbb{E} [W_i Y_i | D_i = 0] &= \mathbb{E} [W_i \mathbb{E} [Y_i(0) | X_i] | D_i = 0] + \mathbb{E} [\eta_i | D_i = 0] \\
&= \mathbb{E} [\mathbb{E} [Y_i(0) | X_i] | D_i = 1] + 0 \\
&= \mathbb{E} [Y_i(0) | D_i = 1],
\end{aligned}$$

car si le score de propension est logit, on peut montrer que  $\mathbb{E} [W_i f(X_i) | D_i = 0] = \mathbb{E} [f(X_i) | D_i = 1]$  pour toute fonction mesurable  $f$ .

On propose l'estimateur:

$$\hat{\theta} = \frac{1}{\sum_{i=1}^n D_i} \sum_{i=1}^n \left( D_i - (1 - D_i) \exp(X_i' \hat{\beta}) \right) Y_i,$$

ce qui correspond à la contrepartie empirique de la condition de moment  $\mathbb{E}[m(Y_i, D_i, X_i, \beta_0, \theta_0)] = 0$  pour  $m(Y_i, D_i, X_i, \beta, \theta) = (D_i - (1 - D_i) \exp(X_i' \beta)) Y_i - D_i \theta$ . Il convient de noter que  $\hat{\theta}$  dépend de  $\hat{\beta}$  qui est une variable aléatoire dont la variance sera à prendre en compte. Une réponse faisant figurer  $\beta_0$  à la place de  $\hat{\beta}$  n'est pas recevable: ce n'est pas un estimateur dans ce cas.  $\hat{\theta}$  est convergent et asymptotiquement Normal, d'après le théorème 6.1 dans

“Large Sample Estimation and Hypothesis Testing” de Newey et McFadden (*Handbook of Econometrics*, 1994) en référence dans l’énoncé, car on se trouve en présence d’un estimateur GMM en deux étapes. On pouvait se contenter d’appliquer le théorème, en considérant que  $(\hat{\theta}, \hat{\beta})'$  est un estimateur GMM tel que :

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n m(Y_i, D_i, X_i, \hat{\beta}, \hat{\theta}) &= 0 \\ \frac{1}{n} \sum_{i=1}^n g(Y_i, D_i, X_i, \hat{\beta}) &= 0.\end{aligned}$$

Le théorème 6.1 donne après calcul :

$$\sqrt{n} (\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\mathbb{E} \left[ \left( (D_1 - (1 - D_1) \exp(X_1' \hat{\beta}_0)) (Y_1 - X_1' \mu_0) - \theta D_1 \right)^2 \right]}{\mathbb{E} D_1} \right),$$

pour  $\mu_0 := \mathbb{E} \left[ (1 - D_1) \exp(X_1' \hat{\beta}_0) X_1 X_1' \right]^{-1} \mathbb{E} \left[ (1 - D_1) \exp(X_1' \hat{\beta}_0) X_1 Y_1 \right]$  et  $\hat{\beta}$  solution de la contre-partie empirique de  $\mathbb{E}[g(D_i, X_i, \beta_0)] = 0$ . On pouvait également appliquer les résultats du cours sur les estimateurs GMM, mais le calcul était plus long.

**Remarque sur la pondération (Partie 3)** Au début de la partie 3, on demande de pondérer les résultats de l’estimation. Cela signifie que plutôt que de minimiser :

$$\frac{1}{n} \sum_{i=1}^n (Y_i - X_i' \beta)^2,$$

où chaque observation a le même poids  $(1/n)$ , on va minimiser :

$$\sum_{i=1}^n \omega_i (Y_i - X_i' \beta)^2,$$

où  $\omega_i$  est le poids associé à l’observation  $i$ . Dans notre application, cela ne fait pas sens de donner le même poids à chaque ville car elles comptent toutes un nombre d’habitants et d’électeurs très différents. On souhaite donc pondérer par le nombre d’électeurs dans la mesure où les résultats permettront de mieux refléter l’impact pour l’électeur moyen.