

# Séries temporelles : dynamique d'adhésions à un parti politique

---

## Contexte

---

La majorité des partis politiques français restent généralement discrets sur leur nombre d'adhérents. Ce n'est pas le cas toutefois de tous les partis. L'objet de cette note est d'étudier en détail l'évolution du nombre d'adhérents quotidien d'un mouvement qui a accepté de donner accès à sa base d'adhérents depuis sa création. Il s'agit de l'U.P.R. (Union Populaire Républicaine), fondée le 25 mars 2007.

## Méthodologie

---

### Obtention des données

Les données nous ont été envoyées après une demande de contact par internet ; la donnée reçue est la combinaison de deux phénomènes (adhésions et radiations), et agrège à la fois les adhésions par carte bleue (heure exacte) et par chèque (comptabilisées à réception ou quelques heures plus tard), ces dernières ne représentant pas plus de 10% du total. L'idéal aurait été de pouvoir accéder à deux séries temporelles différentes : celle des adhésions et celle des radiations (les statuts précisent qu'une radiation intervient 3 ans après la dernière cotisation ou sur demande écrite de l'adhérent), on fera néanmoins l'hypothèse que les radiations sont faibles devant les adhésions, et stables dans le temps.

Par ailleurs nous disposons d'un échantillon de données à pas de temps plus précis (**horaire**) qui sont collectées directement à partir du compteur public du site internet, permettant d'analyser la dynamique au sein d'une même journée.

### Description des données, corrections

La donnée journalière est structurée en deux variables : un horodatage au format JJ-MM-AAAA et un numérique de cumul du nombre d'adhérents chaque jour à 23:00:00. Nous ajoutons alors une variable « adhérents » calculée comme la différence entre le cumul des jours J et J-1 (et par convention nulle au premier jour), ainsi que des variables calendaires descriptives (jour, semaine, mois, etc.). Le très faible nombre de journées négatives semble valider l'hypothèse de radiations faibles et stables.

Tableau 1: Résumé des variables, donnée journalière

	<i>Date</i>	<i>Cumul</i>	<i>Adhérents</i>
<b>Min</b>	25-03-2007	28	-1
<b>Max</b>	30-01-2017	15359	114
<b>N</b>		3600	
<b>Manquants</b>		1	

La donnée journalière est continue dans le temps à l'exception d'une seule date manquante (17-09-2014) dont les adhésions ont été réparties équitablement entre celles de la veille et du lendemain. Trois dates seulement présentent un nombre d'adhésions négatif (-1) qui a été corrigé à zéro pour faciliter la mise en œuvre des algorithmes. Ces corrections peuvent être considérées comme mineures.

Tableau 2: Résumé des variables, donnée horaire

	<i>Date</i>	<i>Cumul</i>	<i>Adhérents</i>
<b>Min</b>	01-01-2016	10104	-1
<b>Max</b>	22-05-2016	11089	16
<b>N</b>		3421	
<b>Manquants</b>		0	

La donnée horaire (Tableau 2) est complète et comporte 11 valeurs d'adhésions négatives (-1) correspondant à un solde négatif *hésions – radiations*. Elles sont conservées.

### Statistiques descriptives journalières

#### Forme générale

La donnée présente les attributs d'une série temporelle avec un caractère non-stationnaire caractéristique à la fois par l'augmentation de sa valeur moyenne dans le temps et de sa dispersion. On la représente graphiquement pour se donner une idée (Figure 1).

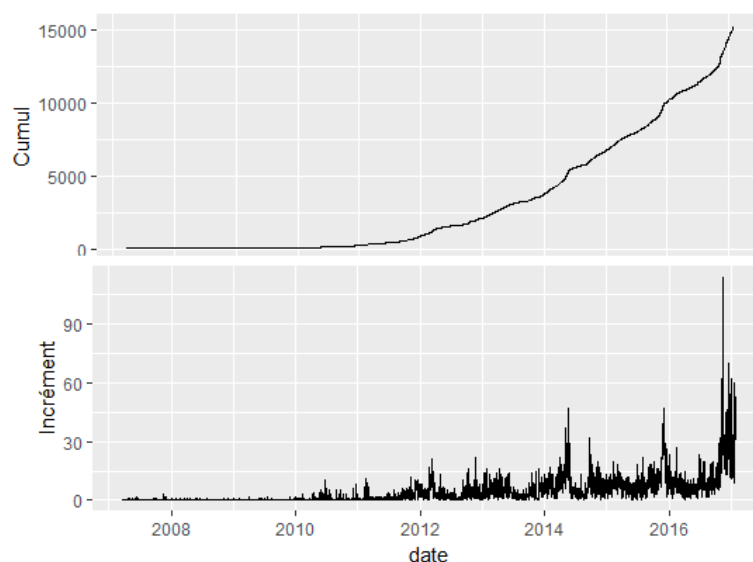


Figure 1: Evolution de la variable « cumul » et « adhérents » dans le temps

## Adhésions : analyse des extrema

Les journées nulles en adhésions se retrouvent naturellement dans les premières années (zones jaunes de la représentation en Figure 2). On dénombre 41% de valeurs nulles.

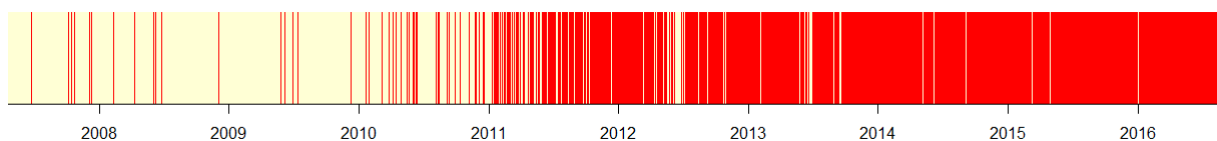


Figure 2: Journées d'adhésions nulles (jaune) ou non-nulles (rouge)

Les 3 derniers événements de ce type sont situés les 4 et 8 avril 2016, et le 26 juillet 2015. Globalement, sur les 10 ans de la série temporelle, les 5 premières années regroupent 95% des journées sans adhésions.

Au contraire, les journées très élevées en adhésions se situent plutôt autour des dates de grandes élections: européennes de 2014 (46 le 22 mai), régionales de 2015 (47 le 2 décembre), et américaines de 2016 (114 le 11 novembre). Ces valeurs sont donc tout à fait valides.

On peut porter au crédit de ce mouvement le maintien d'une croissance soutenue dans le temps. Ainsi quelques chiffres sont parlants :

- Il a fallu 5 ans pour accumuler les premiers 10% d'adhérents.
- Les derniers 21 mois ont vu autant d'adhésions que pendant toute la durée passée.
- Près de 15% des adhérents sont arrivés sur les deux derniers mois et demi.

Le nombre d'adhésions moyen sur toute la période est de 4,26/jour.

## Dynamique de la série temporelle

### Autocorrélation

La série temporelle des adhésions quotidiennes peut être analysée par la fonction d'autocorrélation (Figure 3), qui renseigne sur une éventuelle corrélation entre chaque point et ses précédents.

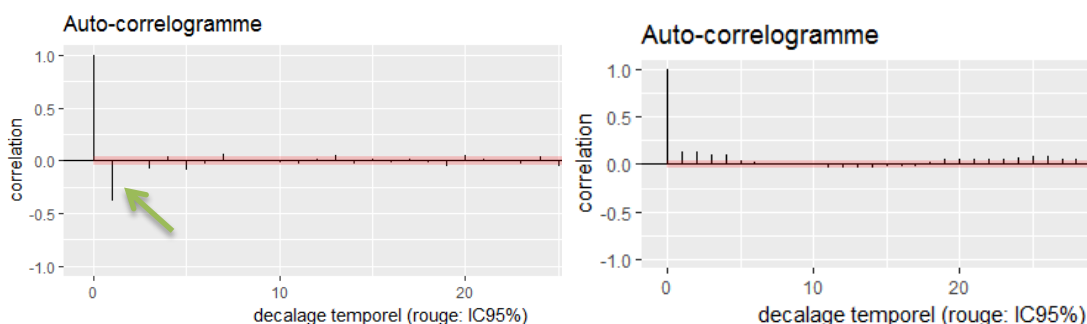


Figure 3: fonction d'autocorrélation pour les adhésions journalières (à gauche), et horaires (droite)

On peut constater que le nombre d'adhésions quotidiennes est corrélé négativement à celui de la veille (flèche verte). Les autres corrélations ne sont pas significatives. On a donc une légère tendance à une alternance entre un jour un peu plus élevé, et un jour un peu plus faible, pouvant refléter des jours avec ou sans publications d'articles motivant l'adhésion. Il n'y a donc pas de « saisonnalité » particulière, comme par exemple des régularités hebdomadaires ou mensuelles. Au pas de temps

horaire, le nombre d'adhésions n'est pas aléatoire mais évolue graduellement au fil de la journée (autocorrélation positive).

### *Dynamique horaire des adhésions*

Le nombre d'adhésions reçues à chaque heure de la journée est représenté dans le graphique suivant. Il n'y a pas lieu de distinguer la semaine et le week-end (un test de modèle linéaire ne conserve pas cette variable). Les adhésions se concentrent sur l'après-midi et la soirée (rythme de 10/jour au lieu de 3/jour sur la nuit et le matin. La courbe du bas rappelle fortement la **fréquentation horaire des réseaux sociaux** (voir le nombre de retweets par heure sur cet [article](#)).

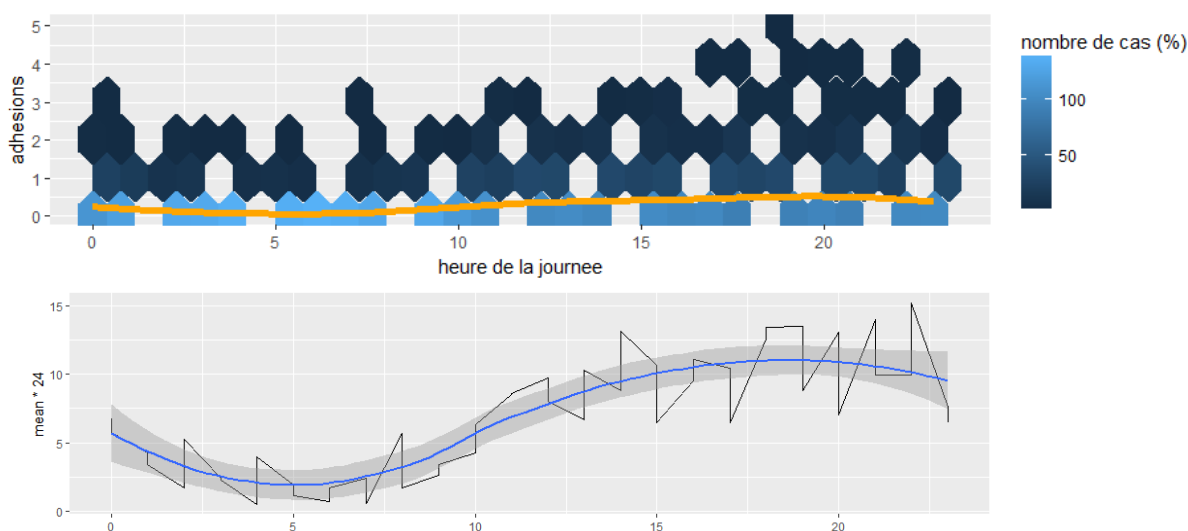


Figure 4: Adhésions par heure de la journée

Sur une période de temps suffisamment faible au regard de la tendance de fond à l'accroissement ( $\leq 3$  mois par exemple), le nombre d'adhésions horaire suit une loi de **Poisson**, dont le paramètre  $\lambda$  varie diffère l'heure de la journée (une loi **exponentielle** est aussi adaptée).

### *Saisonnalité mensuelle*

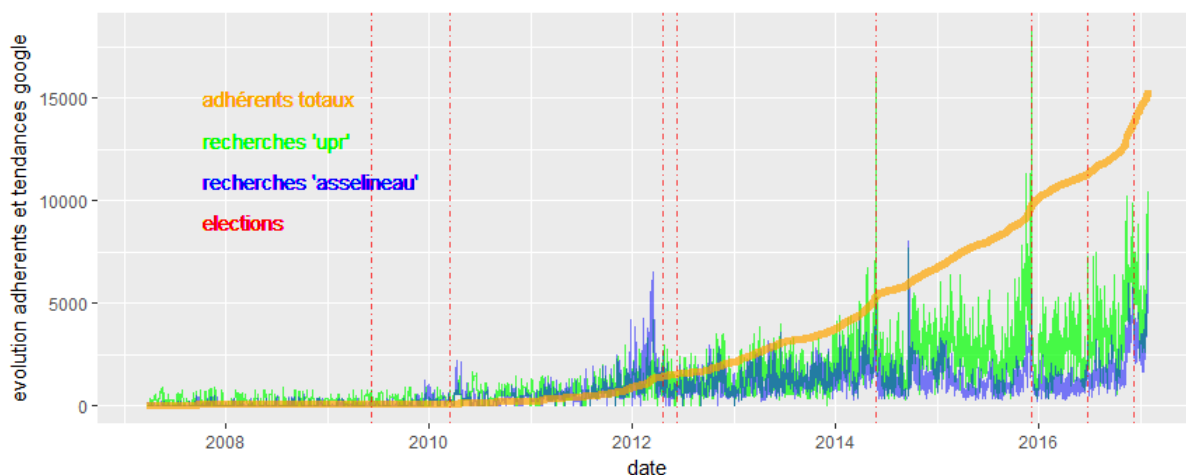
La série des adhésions journalières montre un effet très significatif avec la saison « été » (ainsi qu'une légère autocorrélation annuelle), qui entre début juin et fin août en moyenne accuse une dynamique plus faible (moitié moins d'adhésions que sur le reste de l'année). Une partie de cet écart vient également du fait que les campagnes électorales, sources de notoriété, n'ont pas lieu en été.

### *Corrélation avec des facteurs externes*

Afin de déterminer si la dynamique d'adhésion journalière peut être corrélée à une notoriété médiatique, nous avons récupéré les données de notoriété de mots-clés sur le site [Google Trends](#).

Les données de Google Trends sont en accès libre sous quota, pour n'importe quel mot clé suffisamment requêté depuis janvier 2004, à une échelle journalière. Du fait de la contamination par d'éventuelles recherches homonymes dans d'autres endroits du monde, on se limite aux recherches sur la zone France (les recherches par des expatriés sont donc exclues, ce qui n'est pas anodin car 6% des adhérents seraient des expatriés). Les données sont téléchargées par le biais du package R « [gtrendsR](#) » pour deux mots-clés : « *upr* » et « *asselineau* », ce dernier étant le nom du fondateur et principale figure médiatique du mouvement.

La corrélation entre les recherches « *upr* » et « *asselineau* » sont bonnes (0.65). La corrélation avec le nombre d'adhérents est légèrement meilleure pour « *upr* » (0.67) que pour « *asselineau* » (0.64). Le deux termes ont des volumes de recherche semblables pendant plusieurs années, illustrant le fait que le mouvement politique s'est fondé en identification à son fondateur et principal conférencier ; c'est encore plus flagrant lors de l'élection présidentielle de 2012 qui voit la notoriété de « *asselineau* » largement dépasser celle de « *upr* », ce qui n'arrivera plus jamais par la suite, signe que le parti lui-même devient prépondérant sur son dirigeant (ce qui est bon signe pour sa pérennité). A ce titre, le type d'élection n'est pas anodin puisque déjà la pré-campagne présidentielle 2017 voit le volume « *asselineau* » progressivement rattraper « *upr* », alors que les élections plus locales focalisaient les recherches sur le parti et non sur son dirigeant.

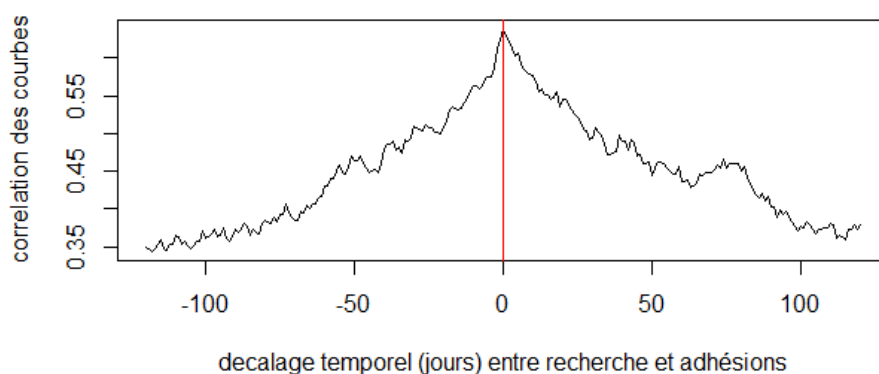


**Figure 5: évolution des adhésions (échelle à gauche) et des recherches Google (échelle relative) et élections (pointillé rouge) : EU(2009), REG(2010), PRES+LEG(2012), EU(2014), REG(2015), BREXIT(2016)\*, USA(2016)\***

**Note :** on observe un pic ponctuel de recherche « *asselineau* » en date de fin septembre 2014 ; c'est la date des élections sénatoriales (mais le mouvement n'a pas été impliqué), c'est également la date à laquelle la page Wikipédia de François Asselineau a été définitivement créée, ce qui peut être l'explication à ce pic très localisé. \*Les élections Brexit et USA2016 sont représentées car ce sont des sujets régulièrement évoqués par ce mouvement.

Une question qu'on peut se poser est : existe-t-il un délai entre la recherche sur Google et la dynamique d'adhésions ? Le graphique suivant (

Figure 6) présente la corrélation entre les deux en fonction du décalage en jours que l'on applique aux trends Google : la meilleure corrélation est observée pour 0 jours de décalage, donc on a un match entre les deux courbes *originales*. Cela peut signifier que la majorité des adhésions ne proviennent pas de gens qui découvrirait soudain ce parti politique, mais plutôt surtout de gens connaissant déjà depuis un certain temps et à qui l'actualité du moment donnerait une raison de faire le pas. Les résultats sont identiques pour le terme « *upr* ».



**Figure 6: décalage temporel entre recherche « *asselineau* » et adhésion?**

### Forme de modèles et signification

Plusieurs techniques de modélisation peuvent être utilisées, comme un modèle linéaire une croissance quadratique ou encore exponentielle :

1. Une **croissance linéaire** modélise un phénomène qui s'accroît d'une quantité fixe à chaque pas de temps, simulant par exemple une situation où les personnes découvrent par hasard le mouvement, à un rythme constant. A chaque nouvel instant, il y a un nombre identique d'arrivants, aux aléas près.
2. La **croissance exponentielle** (ou plus généralement géométrique) incrémente d'une valeur qui augmente elle-même en proportion de la taille de la population. Elle modélise par exemple une situation où chaque nouvel adhérent va convaincre d'autres personnes, dont une partie vont elles-mêmes convaincre d'autre personnes, et ainsi de suite. Tout ceci forme des croissances exponentielles. Ce type d'évolution est très répandu en biologie pour des phénomènes qui démarrent lentement puis finissent par « exploser », mais les croissances exponentielles sont en réalité vite plafonnées par des contraintes physiques du milieu, de l'espace, des interactions, etc., et peuvent évoluer en courbes sigmoïdes (avec point d'inflexion puis plateau) ou cycliques.
3. Une **croissance quadratique** (ou plus généralement polynomiale), c'est-à-dire une parabole (ex :  $y = x^2$ ), possède un accroissement linéaire. Si on double le temps d'attente, on quadruple alors le nombre d'adhérents. Ce type de croissance est rarement rencontré en biologie, on pourrait imaginer ici qu'il reflète plutôt des recrutements proportionnels à une médiatisation qui, elle, croîtrait linéairement.

### Calibration sur les données

Lors des cinq premières années d'existence du mouvement, alors qu'il était encore très confidentiel, le modèle exponentiel (courbe rouge) rend parfaitement compte de la croissance, majoritairement par bouche à oreille puisqu'il n'y a aucune candidature aux élections sur cette période.

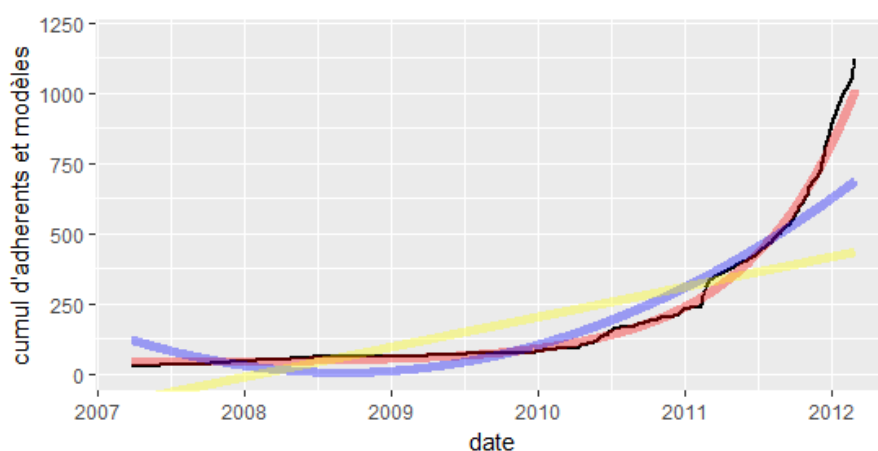


Figure 7: 3 modèles calés sur la période 2007:2012

A partir de 2013, cette première courbe exponentielle ne colle plus aux données: elle croît trop vite. Le rythme du bouche-à-oreille n'augmente plus aussi vite, peut-être que les cercles de relations font que les adhérents prêchent déjà des convaincus.

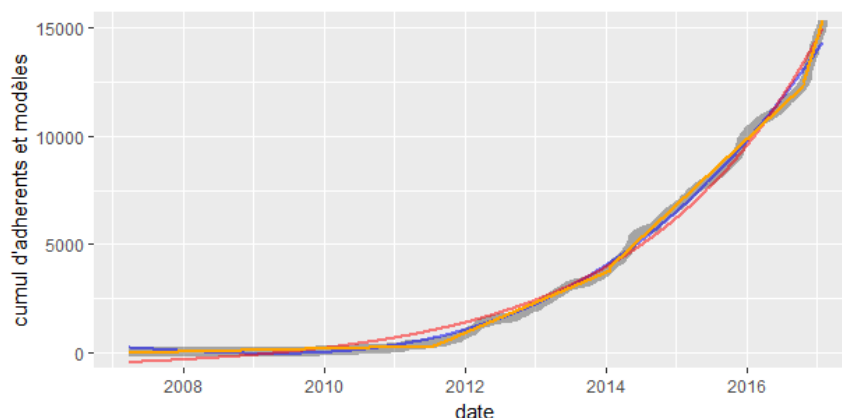


Figure 8: modèles polynomial (deg.3, bleu) et exponentiel (rouge) calés sur la période 2007:2016, et régression linéaire par morceaux (jaune).

La figure ci-dessus (Figure 8) cale ces mêmes modèles sur l'ensemble des données. La meilleure performance ici est le modèle polynomial en  $x^3$ . Ces modèles échouent tous à expliquer les pics d'adhésions des élections ainsi que la forte croissance observée à partir des élections US (fin 2016). Ce dernier phénomène pourrait être lié à une intense campagne d'affichage sur toute la France. Nous avons rajouté sur la figure un modèle linéaire dit « par morceaux » qui cale 4 segments de droite sur la courbe, il obtient un excellent fit (Tableau 3) mais est totalement inutile à la prédiction puisque le choix des points de cassure reste manuel et a posteriori.

Tableau 3: performance de deux modèles (erreur moyenne absolue sur données d'apprentissage)

	Modèle polynomial, 2	Modèle polynomial, 3	Modèle exponentiel	Modèle « segments »
Erreur MAE	287	151	325	85

En conclusion, on peut prétendre expliquer la période après 2013 par un modèle quadratique ou cubique (puissance 3) qui sous-entendrait que la dynamique d'adhésion s'accélère linéairement (ou quadratiquement). La médiatisation croissante peut être une explication, ainsi que la présence sur internet via les réseaux, les commentaires, et les vidéos YouTube. On peut également supposer que le référencement Google du site [www.upr.fr](http://www.upr.fr) (qui est l'un des partis les plus fréquentés d'après le classement AlexaRanking) joue un rôle important. Le référencement naturel Google ramène en tête de recherche en fonction du nombre d'hyperliens et d'interactions entre les sites, ce qui peut faire penser à une loi de type puissance (cf. la [loi de Metcalfe](#) qui établit que la valeur d'un réseau est une fonction quadratique du nombre d'utilisateurs).

### Impact des évènements de type élections

Certaines techniques avancées de traitement des séries temporelles permettent d'estimer l'impact diffus d'un évènement ponctuel sur la série. Ici (Figure 9), nous avons utilisé une méthode plus « empirique » tout d'abord en **lissant** la série temporelle des adhésions par une *smoothing spline* (courbe noire), puis en estimant la **tendance basse** des adhésions journalières (30<sup>ème</sup> percentile) par une régression spline sur quantile (courbe bleue).

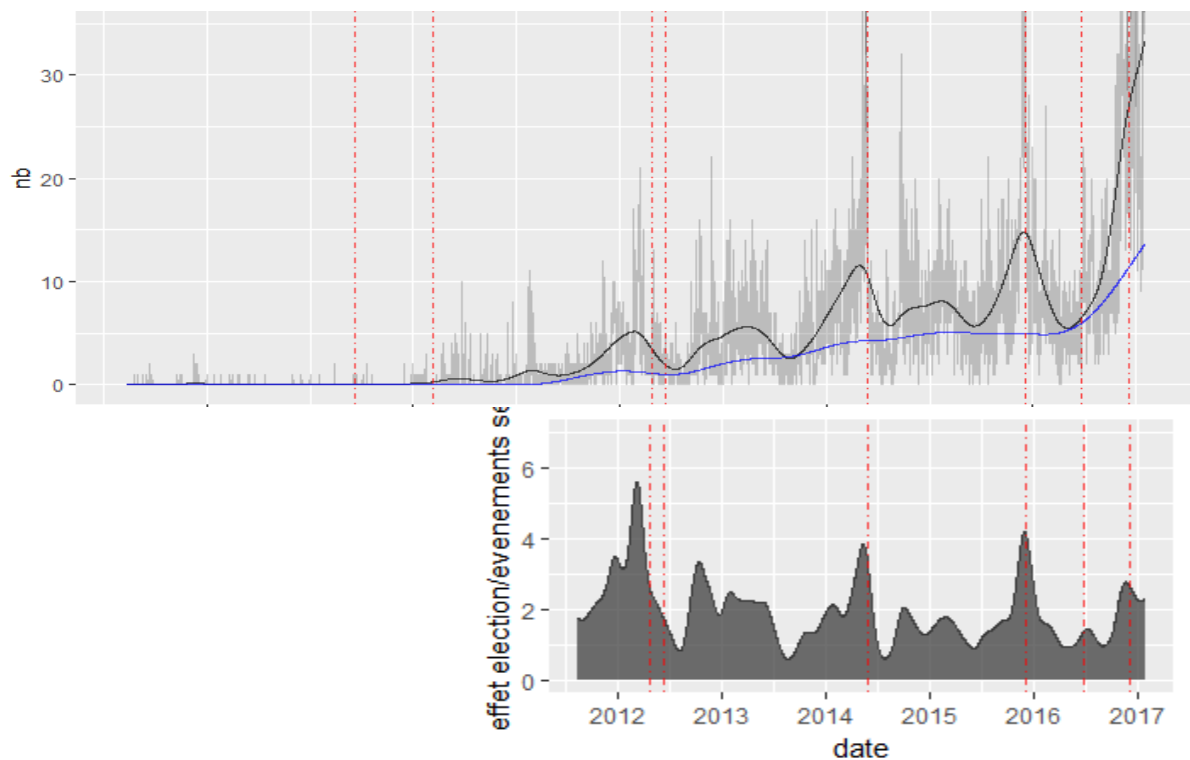


Figure 9: lissage du nombre d'adhésions et facteur multiplicatif électoral

Le ratio entre les deux courbes lissées (figure inférieure) donne une indication du facteur d'augmentation des adhésions dues uniquement au contexte électoral ou événements équivalents, c'est-à-dire un facteur multiplicatif par rapport à un niveau basal hypothétique d'adhésions. L'effet électoral commence souvent plusieurs mois avant l'élection, et peut prendre la forme d'une gaussienne (par exemple pour les élections régionales de 2015, donc un effet symétrique avant et après la date), mais peut aussi s'éteindre quasiment dès le lendemain de l'élection (c'est le cas des européennes de 2014, et en partie de la présidentielle de 2012). D'un point de vue quantitatif, le facteur multiplicatif peut aller dépasser 3 fois le niveau basal d'adhésions sur un créneau d'un mois autour de l'élection.

### Synthèse

---

Tout au long de cette étude, nous nous sommes efforcés d'analyser la dynamique des adhésions à un mouvement politique tout en avançant des hypothèses explicatives, liées aux spécificités de ce mouvement politique. Nous avons modélisé l'évolution des adhésions sur presque 10 années, en décomposant en plusieurs périodes : une phase de croissance exponentielle, puis quadratique, puis enfin une croissance forte sur les derniers mois. Nous avons pu expliquer une partie de ces évolutions par des facteurs comme :

- l'heure de la journée,
- la saison de l'année
- la médiatisation (tendance Google)
- les événements particuliers (élections)

Dans les mois à venir, le suivi de la campagne présidentielle et législative 2017 apportera sans doute des éléments intéressants pour compléter cette étude. L'ensemble des données et scripts statistiques sera bientôt publié sur GitLab (page personnelle [ici](#)). Je tiens enfin à remercier le personnel du mouvement politique UPR pour la fourniture des données ainsi que les précisions par téléphone.