

Données utilisateurs, vie privée et concurrence sur les marchés en ligne

Personnalisation, vie privée et effacement des données

Guillaume Thébaudin
Sous la direction de Marc Bourreau

Doctoriales i^3 , 27 mai 2020

Objectifs de la thèse

- Aborder par le biais de la modélisation théorique économique les problématiques liées à la vie privée en ligne
- Étudier la nécessité d'une régulation (e.g. RGPD)
- Analyser les effets d'une telle régulation sur les dynamiques concurrentielles sur ces marchés.
- Littérature en 'economics of privacy':
 - Casadesus-Masanell and Hervas-Drane (2015): Les entreprises se différencient par le degré de vie privée qu'elles offrent aux utilisateurs.
 - Wynne Lam and Liu (2020): Analyse de l'effet de l'introduction de la portabilité des données sur l'entrée de concurrents dans un marché.
 - Lefouilli and Toh (2020): Analyse l'effet d'une régulation sur la monétisation des données sur le bien-être social, lorsqu'un monopole choisit le niveau de qualité qu'il offre aux utilisateurs
 - Ichihashi (2020): Modèle en dynamique qui montre que sur le long terme, les utilisateurs se comportent comme s'ils n'avaient pas de préoccupations concernant le respect de leur vie privée.
- Premier chapitre: Personnalisation, vie privée et effacement des données.

- Montant croissant de données à caractère personnel généré par l'utilisation de services en ligne.
- Permet aux entreprises de délivrer une expérience utilisateur de plus en plus personnalisée, qui augmente l'utilité que les utilisateurs tirent de l'utilisation des services.
- Exemple de Netflix:
"Personalization is one of the pillars of Netflix because it allows each member to have a different view of our content that adapts to their interests and can help expand their interests over time. It enables us to not have just one Netflix product but hundreds of millions of products: one for each member profile."
- Business model des services en ligne principalement basé sur la monétisation des données: 'Data as a currency'. Les utilisateurs souffrent donc d'une perte de vie privée lorsque leurs données personnelles sont partagées avec des parties tierces (ex: annonceurs publicitaires, data brokers...).

- Lorsque la perte de vie privée des consommateurs liée aux données qu'ils ont délivré par le passé excède les bénéfices d'une utilisation personnalisée du service, les consommateurs souhaiteraient pouvoir demander aux entreprises d'en supprimer.
- Pendant longtemps, les entreprises ne permettaient pas la suppression de données. Récemment, Google a introduit un outils qui permet aux utilisateurs la suppression automatique leur données après une certaine période. Amazon rend également facile la suppression de données enregistrées par ses assistants vocaux.
- Quel montant optimal de données à sauvegarder du point de vue des consommateurs ? des entreprises? Quels facteurs influencent le potentiel désalignement en terme de montant des données à sauvegarder/supprimer ?

- Modèle sur deux périodes: un consommateur choisit son utilisation du service en première période u_1 , ce qui génère un montant u_1 de données permettant à l'entreprise de lui fournir un service personnalisé en deuxième période.
- Entre les deux périodes, le consommateur peut supprimer des données qu'il a généré en première période (d'un montant δ) à un cout unitaire c choisit par la firme, reflétant la facilité d'effacement des données offerte par la firme.
- La désutilité du consommateur liée au fait que ses données soient enregistrées sur les serveurs de la firme est capturé par θ qui reflète le niveau de vie privée que l'entreprise offre aux utilisateurs.
- Utilité du consommateur:

$$U = \underbrace{v(u_1) - \theta u_1}_{\text{Première période}} + \gamma^c \underbrace{(v(u_2) + \lambda(u_1 - \delta)u_2 - \theta(u_1 - \delta + u_2) - c\delta)}_{\text{Deuxième période}}$$

with $v'(\cdot) \geq 0$, $v''(\cdot) = -\phi < 0$, $\lambda'(\cdot) \geq 0$ and $\lambda''(\cdot) = -\eta < 0$

- Profit de la firme, qui maximise le montant des données en qu'elle possède:

$$\Pi = u_1 + \mu(u_1 - \delta) + u_2.$$

Comportement optimal du consommateur lorsque $c = 0$

- Raisonnement à rebours.
- Si on note s_1 le montant de données de première période enregistré $s_1 = u_1 - \delta$, l'utilité du consommateur à cette période est

$$U_2 = v(u_2) + \lambda(s_1)u_2 - \theta(s_1 + u_2).$$

Le montant optimal d'utilisation de seconde période $\tilde{u}_2(s_1)$ est donnée par:

$$v'(\tilde{u}_2(s_1)) + \lambda(s_1) = \theta. \text{ On a: } \frac{\partial \tilde{u}_2}{\partial s_1}(s_1) = \frac{\lambda'(s_1)}{\phi} > 0.$$

- Le montant optimal de données à conserver s_1^* est donné par: $\lambda'(s_1^*)\tilde{u}_2(s_1^*) = \theta$.
- Étant donné la contrainte que $s_1 \leq u_1$, on a

$$s_1(u_1) = \begin{cases} u_1 & \text{if } u_1 \leq s_1^* \\ s_1^* & \text{if } u_1 > s_1^* \end{cases}$$

- Choix de u_1 :

$$v(u_1) - \theta u_1 + \gamma^c \left(v(\tilde{u}_2(s_1(u_1))) + \lambda(s_1(u_1))\tilde{u}_2(s_1(u_1)) - \theta(s_1(u_1) + \tilde{u}_2(s_1(u_1))) \right)$$

$$\begin{cases} v'(u_1^*) + \gamma^c \lambda'(u_1^*)\tilde{u}_2(s_1(u_1^*)) = \theta(1 + \gamma^c) & \text{if } u_1^* \leq s_1^* \\ v'(u_1^*) = \theta & \text{if } u_1^* > s_1^* \end{cases}$$

Suppression couteuse des données

- Hypothèse: lorsque $c = 0$, le consommateur choisit de supprimer un montant positif de données: $\delta^* = u_1^* - s_1^* > 0$
- Effet d'une baisse du montant de données que le consommateur peut supprimer sur l'utilisation du service de première période: si le consommateur peut supprimer uniquement $\delta < \delta^*$, il choisit $\tilde{u}_1(\delta)$ sous la contrainte de $\tilde{s}_1(\delta) = \tilde{u}_1(\delta) - \delta$:

$$v'(\tilde{u}_1(\delta)) + \gamma^c \lambda'(\tilde{u}_1(\delta) - \delta) \tilde{u}_2(\tilde{u}_1(\delta) - \delta) = \theta(1 + \gamma^c).$$

La différenciation de cette expression par rapport à δ donne

$$-\frac{\partial \tilde{u}_1}{\partial \delta}(\delta) = -\frac{\gamma^c \Lambda(\tilde{u}_1(\delta) - \delta)}{\gamma^c \Lambda(\tilde{u}_1(\delta) - \delta) - \phi} < 0 \quad (1)$$

On note que $-\frac{\partial \tilde{u}_1}{\partial \delta}(\delta) > -1$: le consommateur réduit son utilisation de première période face à une baisse du montant qu'il peut supprimer, mais cette baisse est d'une magnitude moindre que la baisse du montant des données qu'il peut supprimer. Par conséquent, lorsque δ baisse, le montant des données enregistrées augmente.

Suppression couteuse des données

- Lorsque δ baisse, du fait que le montant des données enregistrées augmente, l'utilisation du service en deuxième période augmente du fait d'une personnalisation accrue du service:

$$-\frac{\partial \tilde{u}_2}{\partial \delta}(\delta) = -\frac{\lambda'(\tilde{u}_1(\delta) - \delta)}{\phi} \left(\frac{\partial \tilde{u}_1}{\partial \delta}(\delta) - 1 \right) > 0.$$

- Choix du montant de données à supprimer lorsque $c > 0$:

$$-\lambda'(\tilde{u}_1(\tilde{\delta}(c)) - \tilde{\delta}(c))\tilde{u}_2(\tilde{u}_1(\tilde{\delta}(c)) - \tilde{\delta}(c)) + \theta = c.$$

Cela vaut la peine pour le consommateur de payer c pour supprimer une donnée si la désutilité marginale d'avoir cette donnée enregistrée, composée de la différence entre le cout de vie privée θ et le bénéfice marginal d'avoir cette donnée enregistrée, est plus grande que c .

$$\text{On a: } \frac{\partial \tilde{\delta}}{\partial c}(c) = -\frac{1}{\Lambda(\tilde{u}_1(\tilde{\delta}(c)) - \tilde{\delta}(c)) \left(\frac{\partial \tilde{u}_1}{\partial \delta}(\tilde{\delta}(c)) - 1 \right)} < 0.$$

- Il existe un seuil \bar{c} au dessus duquel il est trop coûteux pour le consommateur de supprimer des données qu'il choisit de ne pas le faire. Ce seuil est déterminé par:
$$-\lambda'(\tilde{u}_1(0))\tilde{u}_2(\tilde{u}_1(0)) + \theta = \bar{c}.$$

Choix de c par la firme

- En fixant c , la firme est capable de guider la consommation inter-temporelle du service: lorsque c décroît (croît), la consommation du service de première période augmente (baisse) et celle de deuxième période baisse (augmente), lié à une personnalisation moins (plus) importante.

$$\Pi(\delta) = \tilde{u}_1(\delta) + \gamma^f (\mu(\tilde{u}_1(\delta) - \delta) + \tilde{u}_2(\delta))$$

On fait l'hypothèse que la fonction de profit est concave (elle admet un maximum).

- Il existe un dé-salignement entre les consommateurs et la firme sur le choix du montant des données à sauvegarder, et donc sur le choix de c : pour la firme, le montant optimal dénoté δ^\diamond est donné par :

$$\frac{\partial \tilde{u}_1}{\partial \delta}(\delta^\diamond) = -\gamma^f \left(\mu \left(\frac{\partial \tilde{u}_1}{\partial \delta}(\delta^\diamond) - 1 \right) + \frac{\partial \tilde{u}_2}{\partial \delta}(\delta^\diamond) \right) \quad (2)$$

alors que le choix optimal du consommateur est donné par

$$\lambda'(u_1^* - \delta^*) \tilde{u}_2(u_1^* - \delta^*) = \theta.$$

Analyse du dés-alignement et prochaines étapes

- La firme va choisir un c élevé si son taux de dépréciation du temps est faible ainsi que si le degré de myopie du consommateur est élevé. Plus le taux de dépréciation des données du point de vue de la firme est faible, plus c sera élevé. En revanche, plus les rendements de la personnalisation du service (η) décroissent rapidement, moins la firme choisit un c élevé.
- Prochaines étapes:
 - étude de l'effet de la concurrence sur le choix de c , un entrant entre avec succès en seconde période s'il parvient à fournir un service non-personnalisé (n'ayant pas de données) procurant une plus grande utilité aux utilisateurs que le service personnalisé de la firme présente initialement.
 - choix de θ
 - autres suggestions?