

LES MARCHÉS PRÉDICTIFS COMME TECHNIQUE DE PRÉVISION

Arthur Charpentier

Maître de conférences, Université de Rennes 1

Les dernières élections présidentielles ont remis en avant l'importance des sondages, utilisés comme outils de prévision, même si les sondeurs s'en défendent. Comme l'ont dit Niels Bohr et Pierre Dac, « la prévision est difficile, surtout lorsqu'elle concerne l'avenir », et des solutions alternatives ont été envisagées. Comme c'est la mode, le big data a été mentionné (deviner les intentions de vote à partir des tweets ou des informations publiées sur une page Facebook), mais les marchés prédictifs (et les sites de paris) ont aussi été largement utilisés. Mais est-il si simple d'utiliser les prix (ou les cotes) sur ces marchés prédictifs pour en déduire des probabilités ? Et peut-on vraiment les utiliser ?

Sur quoi parie-t-on ?

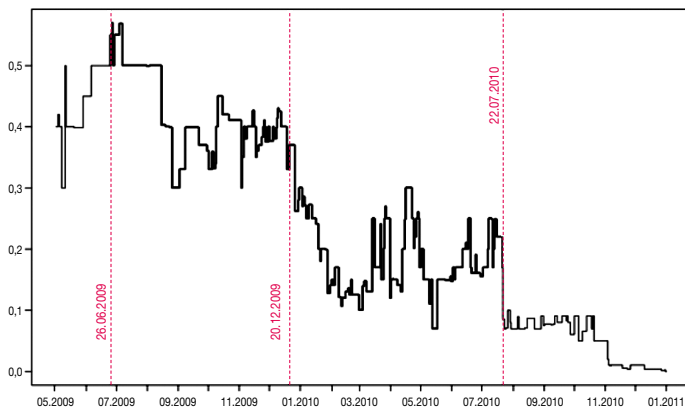
Comme le rappellent Rhode et Strumpf [2008], utiliser les paris pour connaître les croyances et donc les probabilités qu'un événement se réalise n'est pas nouveau ; par exemple lors des élections papales au Vatican. En 1549, Matteo Dandolo (ambassadeur de Vénétie) notait [Baumgartner, 2003] que « *it is therefore more than clear that the merchants are very well informed about the state of the poll, and that the Cardinals' attendants in Conclave (i conclavisti) go partners with them in the wagers, which thus causes many tens of thousands of crowns to change hands* ». Et les marchés de paris lors des élections ont été populaires aux États-Unis jusqu'à la Seconde Guerre mondiale. Rhode et Strumpf [2008] avancent plusieurs raisons pour la perte d'intérêt au cours de la seconde moitié

du XX^e siècle : les améliorations des techniques de sondage et la légalisation des paris sur chevaux (avec pour conséquence la disparition du marché illégal des paris).

Plus récemment, avec le développement des sites et donc d'un marché des paris, on a vu apparaître des paris (contrats vendus sur le marché en ligne www.intrade.com, liquidé en 2015) sur la survenance de tremblements de terre de magnitude dévastatrice (plus de 9 sur l'échelle de Richter), sur le vainqueur de la cérémonie des Oscars, sur la diminution de la couverture de la glace en Arctique entre deux années, ou encore sur l'observation du boson de Higgs. Polgreen *et al.* [2007] évoquent des paris sur la survenance d'épidémies de grippe. Un autre exemple – que l'on va détailler un peu – est celui du passage d'une loi sur l'énergie (et l'instauration de mesures contraignantes de limitation des émissions de gaz à effet de serre) aux États-Unis. Le titre

permettant de toucher un dollar en cas d'adoption du projet de loi a été lancé en mai 2009. Le 26 juin 2009, la Chambre des représentants a adopté (à une courte majorité) le projet de loi (avec pour objectif de réduire de 17 % d'ici à 2020 par rapport à 2005 et de 80 % d'ici à 2050 les émissions de CO₂). À ce moment, le titre était échangé pour cinquante cents comme le montre la figure 1. Il restait à faire approuver la loi par le Sénat. Six mois plus tard, en signant l'accord de la conférence de Copenhague en décembre 2009, les États-Unis s'étaient davantage engagés, et un an plus tard (début 2011), le Sénat serait renouvelé lors des élections de mi-mandat. Le gouvernement de Barack Obama espérait faire approuver la session *lame duck* du Congrès, les élections de mi-mandat devant probablement renforcer la majorité républicaine. Mais le jeudi 22 juillet 2010, le chef de file démocrate au Sénat, Harry Reid, a annoncé le retrait du projet de loi, faute de majorité pour l'approuver. Ce jour-là, le titre s'échangeait pour dix cents.

Figure 1 - Évolution du prix du passage du contrat sur le passage du American Clean Energy and Security Act devant le Sénat (2009-2010)



Source : auteur, données Meng [2017].

Voir un prix comme une probabilité

Avant toutes choses, rappelons qu'une propriété fondamentale des probabilités est que la somme des probabilités d'un événement et de son complémentaire doit être égale à un. C'est assez logique si les probabilités

sont vues comme des « fréquences » correspondant au nombre de cas favorables sur le nombre de cas possibles. Cette interprétation devient plus délicate quand on imagine le cas de variables aléatoires continues, et un passage par la théorie de la mesure s'impose (en particulier pour admettre qu'un « événement de probabilité nulle » puisse malgré tout survenir). Mais historiquement, d'autres définitions des probabilités avaient été proposées, en particulier par Jacques Bernoulli, qui avait introduit une notion de « probabilité morale » (1) que l'on verra réapparaître dans les travaux de Finetti sous le nom de « probabilité subjective », ou à la même époque, dans les travaux de Frank Ramsey [1931]. Pour ce dernier, une probabilité mesure un degré de croyance – *a degree of belief* – qui pouvait se mesurer au travers de paris – *through betting odds* –, et les prix que les joueurs sont disposés à payer donnent une information quant à ces croyances. Il devient alors possible de déduire des probabilités à partir de prix.

L'idée n'était pas nouvelle, puisqu'elle avait été émise dans *Van Rekeningh in Spelen van Geluck* publié par Christiaan Huyghens en 1655 (2) et utilisée quelques années plus tard, en 1671 par Wilhelmina de Witt dans le contexte d'annuités. Le prix d'un contrat versant une rente jusqu'au décès pouvant être vu comme une moyenne pondérée d'annuités (à maturité fixe), en observant les prix des différents contrats d'assurance, on pouvait extraire des probabilités interprétées comme des probabilités de survie. Ces mesures de probabilités n'étaient plus construites à partir de fréquences observées, mais construites à partir de prix historiques observés. On retrouvera cette idée en mathématique financière presque trois siècles plus tard, avec le théorème fondamental de valorisation, et les prix d'Arrow-Debreu. Cette littérature est riche et abondante [Charpentier, 2014], mais il existe également des résultats moins connus sur la création d'un consensus à partir de probabilités et les applications pour les paris.

Travaillant sur le marché du pari hippique, Eisenberg et Gale [1959] ont obtenu des résultats relativement généraux, qui peuvent se formaliser

ainsi : considérons un individu $i \in \{1, 2, \dots, I\}$ qui a la possibilité de parier sur un cheval $j \in \{1, 2, \dots, J\}$ lors d'une course. Chaque joueur possède une somme totale b_i (que l'on normalise de telle sorte que $b_1 + b_2 + \dots + b_I = 1$, et qu'il place la somme $\beta_{i,j}$ sur le cheval j . Ces montants vérifient la contrainte $\beta_{i,1} + \beta_{i,2} + \dots + \beta_{i,J} = b_i$, et on note π_j la somme total placée sur le cheval j , de telle sorte que $\beta_{1,j} + \beta_{2,j} + \dots + \beta_{I,j} = \pi_j$. Compte tenu de la contrainte de budget, notons que $\pi_1 + \pi_2 + \dots + \pi_J = 1$, de telle sorte que les montants π_j peuvent être interprétés comme des probabilités. Supposons maintenant que chaque joueur ait des croyances (que chaque cheval gagne) matérialisées par un vecteur de probabilité $p_i = (p_{i,1}, p_{i,2}, \dots, p_{i,J})$. Eisenberg et Gale [1959] définissent alors une situation d'équilibre, dès lors que si $\beta_{i,j} > 0$, alors

$$p_{i,j} = \pi_j \cdot \max_s \left\{ \frac{p_{i,s}}{\pi_s} \right\}$$

Ils montrent qu'un tel équilibre existe, et qu'il est unique. On retrouve ici un résultat connu en mathématique financière sous le nom de loi du prix unique, même si les croyances ne sont pas explicitement mentionnées. Et c'est ce résultat qui permet d'interpréter les prix sur un marché prédictif comme des probabilités.

En s'inspirant du modèle d'Eisenberg et Gale [1959], Manski [2005] obtient une interprétation sur un marché prédictif. Dans ces marchés, on ne peut pas parier sur un cheval dans une course, mais sur la survenance d'un événement. Considérons un contrat offrant un dollar si un événement A survient (élection d'un nouveau pape, départ d'un pays de la zone euro, etc), proposé au prix π_A . S'il existe un contrat sur l'événement complémentaire \bar{A} (on va supposer ici que c'est le cas), il devrait être proposé au prix

$\pi_{\bar{A}} = 1 - \pi_A$, sinon un arbitrage serait possible (3). Considérons maintenant un agent i croyant avec probabilité p_i que l'événement A va survenir. Si $p_i > \pi_A$, cette personne a intérêt à investir tout son argent pour acheter son titre, sinon, elle a intérêt à investir dans l'autre titre, associé à \bar{A} . Au niveau agrégé, la demande pour les titres de type A et \bar{A} sera respectivement

$$\frac{1}{\pi_A} \sum b_i P[p_i > \pi_A] \text{ et } \frac{1}{\pi_{\bar{A}}} \sum b_i P[p_i < \pi_{\bar{A}}]$$

et on aura un équilibre si

$$\sum b_i = \frac{1}{\pi_A} \sum b_i P[p_i > \pi_A] = \frac{1}{\pi_{\bar{A}}} \sum b_i P[p_i < \pi_{\bar{A}}]$$

Si on suppose que les croyances (p_i) et les richesses (b_i) sont indépendantes, on peut interpréter les prix π_A et $\pi_{\bar{A}}$ comme des probabilités,

$$\pi_A = P[p_i > \pi_A] \text{ et } \pi_{\bar{A}} = P[p_i < \pi_{\bar{A}}]$$

Wolfers et Zitzewitz [2007] proposent une interprétation différente, plus économique, dans le contexte d'agents ayant une utilité logarithmique.

Marché et pari en ligne

La situation que nous venons de décrire est celle d'un marché parfait, où un équilibre se fait entre les parieurs, et cet équilibre donne des prix interprétables comme des probabilités de gain (on parle éventuellement de « pari mutuel »). Mais la pratique des paris en ligne est souvent différente, car les prix sont proposés par un bookmaker. Comme le nom l'indique, ces personnes – à la base – enregistrent les paiements, mais en pratique, ce sont aussi eux qui fixent les prix. Par exemple lors des dernières élections présidentielles en France, plusieurs sites proposaient de parier sur l'issue du second tour.

Tableau 1 - Cotes lors des élections présidentielles de 2017

	Cotes sur les sites de paris en ligne			Intention de votes des instituts de sondage		
	bet365	comeon	betsafe	Financial Times	Bloomberg	BBC
Emmanuel Macron	1/8	1/8	1/7	64 %	60,5 %	61 %
Marine Le Pen	5/1	9/2	9/2	36 %	39,5 %	39 %

Source : <http://www.onlinebettingsites.com/2017-french-election/> (données fin avril 2017).

Comme le montre le tableau 1, les cotes offertes par les bookmakers sont différentes entre les sites de paris. Une cote de 1/8 signifie qu'en pariant huit euros, mon gain net sera de 1 € si le candidat l'emporte le jour de l'élection (et je perds ma mise si le candidat perd). La « probabilité » associée est ici $\pi_A = 8/9 \approx 88,89\%$ pour E. Macron (mon gain brut étant de 9/8 € pour 1 € misé) et $\pi_{\bar{A}} = 1/6 \approx 16,67\%$. Le souci – on retrouve ici la discussion initiée par Jacques Bernoulli – est que ces « probabilités » ne sont pas égales à 1, puisque $8/9 + 1/6 \approx 1,055\%$. La différence vient du fait qu'il ne s'agit pas d'un « juste » prix, le bookmaker s'assurant un rendement certain (de l'ordre de 5,5 % de la mise initiale, quel que soit le site choisi). Les vraies probabilités induites par ces cotes sont ici respectivement $\pi_A = 16/19 \approx 84,2\%$ et $\pi_{\bar{A}} = 3/19 \approx 15,8\%$. Les sites de paris en ligne donnaient une probabilité de gain pour Emmanuel Macron de 84 %, alors que les sondages indiquaient une intention de vote de 62 % pour les instituts de sondage.

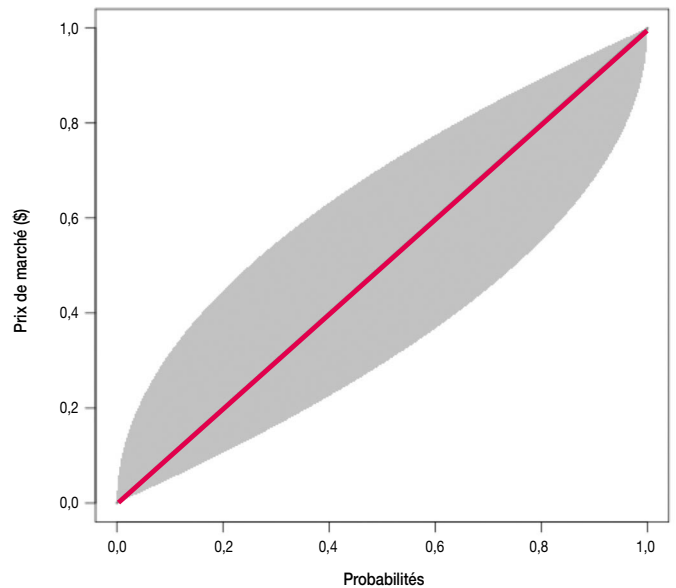
■ Réelle probabilité ou fourchette ?

On le voit, les grandeurs données ne sont pas identiques, l'une indiquant une croyance sur une probabilité de gagner, et l'autre une proportion de votes que devrait avoir un candidat. Dans le modèle de Manski, les probabilités induites par la moyenne des croyances et les prix coïncident si on suppose que les agents ont une utilité logarithmique. Comme le montrent Wolfers et Zitzewitz [2006], si on suppose que les agents ont une aversion au risque, ça ne sera plus le cas. En particulier, l'article suggère de considérer des bornes pour π_A de la forme $[\pi_A^2, 2\pi_A - \pi_A^2]$, représentées sur la figure 2, où π_A est appelée « prix de marché » pour reprendre une notation proche de la figure 1.

■ Faut-il utiliser les prédictions des marchés prédictifs ?

On l'a vu lors de la dernière élection en France, beaucoup s'interrogent sur la place prise par les

Figure 2 - Bornes $[\pi_A^2, 2\pi_A - \pi_A^2]$ sur les croyances moyennes



Source : auteur.

sondages, en particulier en notant qu'il ne s'agissait pas d'une photographie neutre, mais que les électeurs pouvaient choisir leur candidat précisément à la lecture des sondages (la fameuse théorie du « vote utile » en particulier). Le danger à utiliser les marchés prédictifs comme outils de prévision est que ces données sont manipulables, comme l'a montré Schröder [2009]. Un autre danger – et cela a été noté lors de la dernière élection présidentielle aux États-Unis –, les cotes sur les sites en ligne de paris reflètent les croyances d'une certaine élite (disons les personnes souvent relativement éduquées, et la classe supérieure, suivant la politique), contrairement aux sondages qui sont supposés être conduits sur un échantillon représentatif des électeurs, et sont supposés refléter les choix du « peuple » (et non pas une opinion ou une croyance). Si les parieurs tiennent compte de ces sondages, les prix reflètent leurs croyances. Une dernière critique que l'on peut faire est que la majorité des marchés prédictifs représentent des volumes de transactions très faibles.

Les sondages traditionnels font face aujourd'hui à une défiance croissante. Il devient difficile de constituer un échantillon représentatif, le taux de non-réponse atteignant des proportions qui n'avaient encore

jamais été constatées (et nul ne sait si cette non-réponse est corrélée avec des choix électoraux). Et la confusion règne entre les sondages classiques et les sondages en ligne hébergés sur divers sites d'information. Mais pour reprendre l'anecdote racontée par Scott Armstrong dans Knowledge@Wharton en 2004, si dans une réunion du conseil exécutif d'une grande entreprise on demande à tous les participants de faire des prévisions quant aux résultats pour l'année prochaine, ou si on demande aux participants de parier sur les résultats de l'année prochaine, quelles prévisions seront les plus crédibles ?

Notes

1. Comme le rappelle Johnson [2017].
2. En latin sous le titre De Ratiociniis in Ludo Aleæ.
3. Comme l'a montré Franck et al. [2012], comme il existe aujourd'hui plusieurs sites de paris (plusieurs marchés), il peut exister des arbitrages entre les marchés, rarement au sein d'un marché.

Bibliographie

- ARROW K.J. ; DEBREU G., "Existence of an Equilibrium for a Competitive Economy", *Econometrica*, vol. 22, 1954.
- BAUMGARTNER F., *Behind Locked Doors: a History of Papal Elections*, Palgrave Macmillan Trade, 2003.
- CHARPENTIER A., « L'efficiency des marchés : hypothèse de modèle ou fait stylisé ? », *Risques*, n° 96. 2014.

EISENBERG ED. ; GALE D., "Consensus of Subjective Probabilities: The Pari-Mutuel Method", *The Annals of Mathematical Statistics*, vol. 30, 1959, n° 1, pp. 165-168.

FRANK E., VERBEEK E. ; NÜESCH S., "Inter-Market Arbitrage in Betting", *Economica*, 2012.

JOHNSON T., *Ethics in Quantitative Finance: a Pragmatic Financial Market Theory*, Wiley, 2017.

Knowledge@Wharton, "How Credible Are Polls? Is There a Better Way to Predict Outcomes in Politics and Business", 3 novembre 2004. <http://knowledge.wharton.upenn.edu/article/how-credible-are-polls-is-there-a-better-way-to-predict-outcomes-in-politics-and-business/>

MANSKI C.F., "Interpreting the Predictions of Prediction Markets", NBER Working Paper, n° 10359, 2004.

MENG K.C., "Using a Free Permit Rule to Forecast the Marginal Abatement Cost of Proposed Climate Policy", *American Economic Review*, vol. 107, 2017.

POLGREEN P.M. ; NELSON F.D. ; NEUMANN G.R., "Use of Prediction Markets to Forecast Infectious Disease Activity", *Clinical Infectious Diseases*, vol. 44, n° 2, janvier 2007, pp. 272-279.

RAMSEY F.P., "Truth and Probability", in *The Foundations of Mathematics and Other Logical Essays*, 1931, réédition 2013, R. B. Braithwaite Ed.

RHODE P.W. ; STRUMPF K., *Historical Election Betting Markets: an International Perspective*, 2008. https://www.unc.edu/~cigar/papers/PoP_submit4.pdf

SCHRÖDER J., "Manipulations in Prediction Markets", Universitätsverlag Karlsruhe, 2009.

WOLFERS J. ; ZITZEWITZ E., "Interpreting Prediction Market Prices as Probabilities", NBER Working Paper, n° 12200, 2006.