

MICHAEL S. LEWIS-BECK

## Comparaison de prévision des élections présidentielles en France et aux États-Unis

*Journal de la société statistique de Paris*, tome 136, n° 1 (1995),  
p. 29-45

[http://www.numdam.org/item?id=JSFS\\_1995\\_\\_136\\_1\\_29\\_0](http://www.numdam.org/item?id=JSFS_1995__136_1_29_0)

© Société de statistique de Paris, 1995, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

# COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES EN FRANCE ET AUX ÉTATS-UNIS

Michael S. LEWIS-BECK  
Département de Science Politique  
Université d'Iowa, États-Unis<sup>1</sup>

## 1. Introduction

En France comme aux États-Unis, les prévisions électorales sont devenues une entreprise de recherche sérieuse. L'intérêt pour les élections présidentielles est des plus vifs, ceci étant sans aucun doute dû à l'importance des enjeux. Dans les pages qui suivent, je vais étudier les tentatives de prévision des élections présidentielles dans chacun des deux pays, en m'intéressant plus particulièrement aux élections les plus récentes. Je vais par ailleurs systématiquement comparer les succès de ces prévisions de manière à essayer de trouver des modèles de base établis sur une "forte théorie". Comme il le sera montré, un modèle de structure simple peut donner des prévisions relativement justes, plus ou moins bonnes, que les élections aient lieu en France ou aux États-Unis. Je m'appliquerai plus particulièrement à ce modèle afin d'essayer de prédire les prochaines élections présidentielles françaises qui auront lieu en 1995.

---

1. Je voudrais tout particulièrement remercier Georges Prat, le Président de la SSP. Je tiens aussi à remercier, pour son support, la Fondation Camargo de la ville de Cassis, en France, où je résidais alors que je travaillais sur ce projet. Enfin, je voudrais exprimer ma gratitude à Virginie Delfosse Reese, étudiante en doctorat de Français à l'Université d'Iowa, pour son excellente traduction du texte original anglais.

## 2. Prévision des élections présidentielles aux États-Unis

La victoire de Bill Clinton aux élections présidentielles de novembre 1992 a marqué la fin d'un certain nombre de règles populaires de prédiction, basées sur le championnat du monde de base-ball, la vengeance du Beaujolais, la taille du candidat et le vote-témoin du comté de Crook, pour ne citer que quelques exemples (Sur le sujet de ces "prévisions à partir de coïncidences", voir Lewis-Beck et Rice, 1992). De la même manière, les représentants des deux camps de prévision scientifique aux États-Unis – les *modèles* et les *sondages* – ne sont pas toujours tombés juste. En ce qui concerne le camp des modèles, l'échec le plus connu est sans doute celle de l'économiste Ray Fair, qui prédit une victoire écrasante de Bush avec 55,7 % des voix de l'ensemble du corps électoral. Après la publication des résultats véritables, il déclara noblement au sujet de son modèle qu'il devrait sans doute "simplement s'en débarrasser" (*New York Times*, le 5 novembre 1992, B1). Naturellement, les sondages peuvent aussi se tromper. Selon un sondage Gallup portant sur des électeurs probables, réalisé les 26 et 27 octobre (c'est-à-dire quelques jours avant les élections), Clinton et Bush étaient engagés dans une course trop serrée pour être décisive (ils étaient alors crédités respectivement de 40 % et de 38 % des voix).

Ce ne sont là que des exemples, mais ils sont les révélateurs d'un problème : des méthodes de prévision efficaces ont prédit des résultats très différents, depuis une victoire écrasante de Bush jusqu'à une victoire confortable de Clinton. Un moyen de résoudre ce problème serait de renoncer à toute prévision et de se limiter à une explication des résultats à la suite des élections. Un autre moyen, que je préfère, serait de trouver les erreurs de méthode des différents modèles. Je m'attacherai d'abord au problème des sondages effectués avant les élections aux États-Unis, avant de m'attacher aux problèmes des modèles statistiques.

### A) Sondages

La plus vieille méthode de prévision scientifique est celle du sondage d'opinion publique. Selon un procédé maintenant familier, on prélève un échantillon national d'adultes représentatifs et on leur demande comment ils voteraient. On pose aux personnes interrogées une question du genre : "Si les élections présidentielles avaient lieu aujourd'hui, est-ce que vous choisiriez X ou Y pour Président (ou Z, comme avec Perot)?" Dans le sondage Gallup mentionné plus haut, en plus des 40 % de Clinton et des 38 % de Bush, il y avait les 16 % de Perot et 6 % de sans opinion. Si l'on admet que le sondage est juste, à l'intérieur des marges d'erreur d'échantillonnage, alors la course est trop serrée pour être décisive. Mais prenons un autre sondage réalisé par l'institut Harris, quatre jours avant les élections (voir Tableau 1). Selon ce sondage, Clinton recevrait 46 % des voix et Bush 32 %, ce qui donnait à Clinton une avance de 14 %.

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

*Tableau 1.* Intentions de vote d'électeurs probables, pour les élections présidentielles, selon un choix de sondages médiatiques effectuées entre les 20 et 27 octobre 1992

Institut de sondage	Clinton	Bush	Perot	Sans opinion	Avance de Clinton
Gallup pour CNN/USA Today *	40 %	38 %	16 %	6 %	+ 2
Harris @	46 %	32 %	18 %	4 %	+ 14
NBC News/Wall Street Journal #	43 %	36 %	15 %	7 %	+ 7
<b>Résultats des élections</b>					
Voix d'ensemble du corps électoral	43 %	38 %	19 %		+ 5
Voix du collègue électoral	370	168			
<i>Source : Le New York Times, 29 octobre 1992, A12.</i>					
<i>Les données des élections sont issues du New York Times du 5 novembre 1992, B2.</i>					
<p><b>Note :</b> * N = 1 217 électeurs probables (erreur d'échantillonnage de <math>\pm 3</math>) interrogés les 26 et 27 octobre ; @ N = 1 021 électeurs probables (erreur d'échantillonnage de <math>\pm 3</math>) interrogés du 20 au 22 octobre ; # N = 576 électeurs probables (erreur d'échantillonnage de <math>\pm 5</math>) interrogés le 27 octobre.</p>					

L'écart entre les deux candidats de tête enregistré par ces deux instituts réputés, 2 points contre 14 points, est trop grand pour pouvoir être attribué à une erreur d'échantillonnage. Qui plus est, ce n'est pas exceptionnel pour des sondages d'intention de vote réalisés avant des élections présidentielles. [En 1984, le dernier sondage pré-électoral donnait à Reagan une avance de 10 à 25 % sur Mandale, selon les instituts de sondage (Lewis-Beck et Rice, 1992, p. 16)]. Par ailleurs, ni Harris ni Gallup ne sont tombés juste en 1992, puisque selon les résultats finaux du vote véritable de l'ensemble du corps électoral, Clinton avait une avance de 5 % (voir Tableau 1). Pour la période juste avant les élections, un sondage qui a, en fait, assez bien estimé l'écart de victoire de Clinton est un sondage réalisé par *NBC News* et le *Wall Street Journal* (voir Tableau 1).

Maintenant réfléchissons à l'impact potentiel de ces données. Si vous lisiez le résultat de ces sondages qui ont tous été publiés dans le *New York Times* paru le matin du 29 octobre 1992, quel serait votre pronostic ? Supposons que vous ayez porté votre choix sur un sondage, celui de Gallup par exemple, en raison de sa réputation scientifique et de ses antécédents. Il serait impossible de déterminer les résultats de la course, étant donnés les critères habituels (c'est-à-dire 95 % d'intervalle

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

de confiance). Perplexe, vous pourriez vous tourner vers un signe “obscur” (les trois sondages placent Clinton en tête), ce qui vous conduirait à faire une prévision en faveur de Clinton, mais probablement sans grande conviction.

Pourquoi ces instruments de mesure apparemment équivalents diffèrent-ils tant l'un de l'autre, et souvent aussi du résultat même des élections ? Il existe plusieurs raisons possibles : la composition du questionnaire, la réaction des personnes interrogées, l'instabilité des électeurs, la répartition des “sans opinion”, l'identification des électeurs mêmes et les problèmes de construction de l'échantillonnage. L'institut qui a le mieux réussi à résoudre ces problèmes, tout au moins de réputation traditionnelle, est l'institut Gallup. Cependant, même dans leur cas, le résultat des prévisions est irrégulier. Examinons les sondages d'intention de vote réalisés par Gallup quelques jours avant les élections présidentielles entre 1948 et 1988 [voir Lewis-Beck et Rice (1992, p. 17)]. Dans deux de ces courses à la présidence, le candidat perdant est celui qui était en tête des sondages (en 1948 et 1976). Par ailleurs, dans seulement six des onze sondages, il était possible de déclarer avec certitude la victoire d'un des candidats, en tenant compte des limites conventionnelles d'erreur d'échantillonnage (en 1952, 1956, 1964, 1972, 1984 et 1988). Naturellement, même le fait que six des onze sondages aient donné des prévisions justes implique l'hypothèse potentiellement problématique que le vote des “sans opinion” s'est réparti essentiellement de la même manière que celui des “avec opinion”.

Supposons, dans l'intérêt de notre discussion, que le sondage d'intention de vote réalisé par Gallup la veille des élections présidentielles, soit un instrument virtuellement incontestable, capable de prédire des résultats contenant seulement un minimum d'erreur. Est-ce que ce serait alors l'instrument de prévision par excellence ? Nous dirions non pour au moins trois raisons. La première raison, la plus importante, est que le sondage n'est basé sur aucune théorie de vote et donc ne peut nous aider à “expliquer” le résultat des élections. La deuxième est qu'il ne fournit aucune avance dans ses prévisions, prévoyant essentiellement ce qui va se passer demain à partir des données d'aujourd'hui. Ainsi, des prévisions exactes faites la veille des élections, telles celles faites par Gallup en 1992, sont insignifiantes. Comme l'ont dit Nelson Polsby et Aaron Wildavsky (1984, p. 206) : “Ce genre de prédiction n'a pas grande importance. Après tout, on apprend assez rapidement qui a gagné”. La troisième raison est que, pour des chercheurs en sciences humaines indépendants, il est très coûteux, en matière de temps et d'argent, d'effectuer des sondages au niveau national. Ces difficultés, pour n'en mentionner que quelques-unes, ont conduit de nombreuses personnes à adopter une différente stratégie de prévision, celle des *modèles statistiques*.

### **B) Les modèles statistiques**

La méthode des modèles employée pour établir des prévisions diffère de manière essentielle de la méthode des sondages. Les partisans du modèle présentent une théorie de résultats électoraux, théorie généralement basée sur un petit nombre de variables explicatives tirées de la littérature sur le comportement électoral. Une variable de base traditionnelle est la performance économique, mais les autres

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

variables tendent à se différencier. Par ailleurs, l'utilisation de variables de conception identiques peut changer grandement. Les modèles sont formulés comme des systèmes à équation unique, qui essaient de prédire le résultat des élections présidentielles. L'estimation est presque toujours les moindres carrés ordinaires (MCO) sur des séries chronologiques nationales agrégées de la période d'après la Seconde Guerre mondiale. Le tableau 2 contient des modèles connus, ainsi que leurs prévisions pour les élections de 1992.

*Tableau 2. Avance et exactitude de différents modèles de prévision des élections présidentielles en 1992*

Modèle	Avance *	Prévision pour Bush en 1992 @	Prévision substantive pour Bush	Erreur de points en 1992 #
Abramowitz	Longue	46,7 %	Défaite manifeste	- 0,2
Brody et Sigelman	Courte	44,8 %	Grande défaite	- 2,1
Campbell et Wink	Courte	49,2 %	Défaite de justesse	+ 2,3
Fair	Courte	55,7 %	Grande victoire	+ 8,8
Hibbs	Courte	48,9 %	Défaite de justesse	+ 2,0
Lewis-Beck et Rice	Longue	51,8 %	Victoire de justesse	+ 4,9
<b>Résultats des élections</b>		46,9 %	Défaite manifeste	

**Note :** \* L'avance indique combien de temps avant les élections les mesures du modèle en question étaient disponibles. "Longue" équivaut à quatre mois ou plus avant les élections. "Courte" équivaut à environ un mois avant les élections. "Aucune" signifie que les mesures n'étaient disponibles qu'après les élections.

@ La part prévue du vote de l'ensemble du corps électoral en faveur de Bush, en supposant que ce vote est distribué entre les deux partis majeurs.

# Les prévisions moins la part véritable des voix.

**Sources :** Les prévisions des modèles de Fair, et Lewis-Beck et Rice (en tant que prévision du vote de l'ensemble du corps électoral) sont publiées dans l'ouvrage de Campbell et Mann (1992, p. 23). La prévision du modèle pour Campbell et Wink est directement calculée à partir de leur formule, présentée dans l'ouvrage de Campbell et Mann (1992, p. 24), en utilisant les données du sondage réalisé par *New York Times/CBS* en septembre et donnant comme résultats : Bush (34), Clinton (42), Perot (14), Sans opinion (11). La prévision d'Abramowitz a été donnée par lui-même à une conférence de l'Association Américaine de Science Politique, en septembre 1992. La prévision de Brody et Sigelman a été calculée à partir de leur équation d'origine (voir Brody et Sigelman, 1983), en utilisant l'évaluation de l'approbation du Président faite par *New York Times/CBS* en octobre 1992 et qui est de 37. La prévision de Hibbs vient d'une note reçue de lui (Hibbs, 6 octobre 1992).

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

Les modèles semblent varier énormément en ce qui concerne la qualité de leur prévision. Lewis-Beck et Rice (1992, p. 95) ont comparé les modèles de manière générale, en terme de la fiabilité de leurs statistiques. Le R-carré varie entre 0,63 et 0,95 et l'écart type d'estimation (ETE) de la variable dépendante de la part des voix de l'ensemble du corps électoral reçues par le candidat sortant, varie entre 1,40 et 5,09. De la même manière, les modèles varient considérablement en matière d'avance : certains d'entre eux (par exemple Lewis-Beck et Rice) donnent des mesures appropriées de variables de prévision quatre ou cinq mois avant les élections, tandis que d'autres (par exemple Tuft) n'ont pas de mesures appropriées disponibles avant que les élections soient terminées.

Pour ce qui est des élections présidentielles de 1992 en particulier, on peut noter l'existence d'un large éventail de prévisions en ce qui concerne la part du vote du corps électoral des deux partis en faveur du candidat sortant, celle-ci allant de 44,8 % jusqu'à 55,7 %. Ceci est un large éventail surtout lorsque l'on considère le *swing-ratio* entre le vote de l'ensemble du corps électoral et le vote du collège électoral. [1 % du vote de l'ensemble du corps électoral équivaut à un peu plus de 4 % du vote du collège électoral ; voir Lewis-Beck et Rice (1992, p. 56).] En substance, cela veut dire que les prévisions allaient d'une victoire confortable (presqu'écrasante) de Clinton à une victoire écrasante de Bush.

Une fois de plus, nous demandons au lecteur de réfléchir de manière prospective. Supposons que vous lisiez ces prévisions (qui étaient à la disposition du public) avant le jour des élections. Si vous ne disposez d'aucune autre information, les différentes prévisions offrent peu d'indication en ce qui concerne l'identité de la personne qui remportera les élections et avec une avance de combien. Cependant, il vous serait peut-être utile de savoir que, parmi tous ces modèles, celui de Campbell et Wink contient le plus de statistiques fiables ; R-carré corrigé = 0,94 et l'écart type d'estimation = 1,52. Leur pronostic de 49,2 % en faveur de Bush anticipe une course très serrée, une course en fait trop serrée pour qu'une victoire décisive puisse être déclarée, compte tenu de l'écart type d'estimation (même un écart type d'estimation de + 1 ou - 1 offre à Bush de véritables chances de succès). Une autre information qui pourrait être utile est l'avance des modèles respectifs. On pourrait par exemple éviter le modèle Abramowitz pour la bonne raison qu'il utilise de vieilles données (datant du milieu de l'été). Cependant, il s'avère que le modèle Abramowitz est le plus juste de tous, ayant prédit la part des voix de l'ensemble du corps électoral en faveur de Bush presque correctement.

Est-ce que les "astucieux" devraient maintenant se fier au modèle Abramowitz pour les élections de 1996 ? Peut-être, mais gardons en mémoire le cas du modèle Fair, qui réussit bien dans le passé, mais eut les pires résultats en 1992 ; ou bien, celui du modèle Hibbs, qui l'emporta sur tous les autres en 1984, mais faisait partie du peloton lors des dernières élections. [Voir Lewis-Beck (1985a).] La plus grande part de l'instabilité des prévisions fournies par les partisans du modèle vient sans aucun doute du problème du petit nombre d'observations. Dans le cas des élections de 1992, la plupart des partisans du modèle ont basé leurs calculs sur  $N = 11$ . Il est évident que dans la mesure où aux États-Unis les élections présidentielles n'ont lieu

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

que tous les quatre ans, le problème posé par le petit nombre d'observations persistera pendant quelque temps. Cependant, en dehors d'une augmentation du nombre d'observations, il existe d'autres moyens susceptibles d'améliorer le résultat des modèles. J'ai notamment à l'esprit des améliorations en matière de précision des modèles. En raison du nombre limité de degré de liberté, nos modèles doivent être minutieux. Par ailleurs, les quelques variables utilisées comme facteurs de prédiction, devraient être basées sur une théorie forte. Dans les lignes qui suivent, je discuterai les erreurs possibles de précision dans la formulation faite par Lewis-Beck et Rice, appelée Modèle II et présentée dans le tableau 2. Je proposerai ensuite une formulation de base plus minutieuse, appelée Modèle I. [Pour un développement complet de ces modèles, voir Lewis-Beck et Rice (1992, chap. 3). La distinction entre le Modèle I et le Modèle II est une question de commodité et a été faite pour la première fois par Campbell et Mann (1992, p. 23).]

### C) Erreur de spécification dans le Modèle II

Théoriquement, le Modèle II dit que les électeurs aux élections présidentielles prennent une décision basée sur des questions économiques et non économiques, le partisanat et les caractéristiques du candidat sortant. En quelques mots, le modèle actuel peut être écrit de la manière suivante :

$$\begin{aligned} \text{le vote présidentiel} = & \text{les conditions économiques} \\ & + \text{la popularité du Président} \\ & + \text{la force du parti} \\ & + \text{l'attrait du candidat} \end{aligned}$$

Si l'on utilise ces variables et l'on se livre à une estimation à l'aide des moindres carrés ordinaires sur des séries chronologiques nationales d'élections ayant eu lieu entre 1948 et 1988, on obtient l'équation suivante :

$$\text{Equation 1} \quad V = 6,83 + 7,76* G + 0,86* P + 0,52* S + 19,66* C$$

(0,50) (3,79)      (3,39)    (2,87)      (3,30)

R-carré = 0,95

R-carré corr. = 0,92

N = 11

ETE = 9,10

D - W = 2,34

où V = la part de vote du collègue électoral (en pourcentage) pour le candidat du parti sortant ; G = le taux de croissance réel du PNB (non annuel) pendant les six premiers mois de l'année de l'élection ; P = l'appréciation du Président (en pourcentage) selon le sondage Gallup du mois de juillet ; S = le nombre total de sièges perdus à la Chambre des Représentants par le parti du Président lors des dernières élections législatives ayant eu lieu à la moitié du mandat présidentiel ; C = le nombre total de votes reçus aux élections primaires (en pourcentage) par le candidat du parti sortant, recevant "1" si le nombre est supérieur à 60 %, et "0" dans tout autre cas ; les chiffres entre parenthèses sont les écart-types ; \* = statistiquement significatif à 5%, le T de Student pour un test unilatéral étant à 1,94 ; R-carré = le coefficient de



## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

détermination multiple ; R-carré corr. = le coefficient de détermination multiple corrigé ; ETE = l'écart type d'estimation ; N = le nombre d'élections présidentielles ayant eu lieu entre 1948 et 1988.

Ce modèle, le Modèle II, a prédit le vainqueur dans dix des onze élections ayant eu lieu pendant la période d'estimation (se trompant seulement en 1960, avec Kennedy). Par ailleurs, l'erreur en dehors de l'échantillon, représentée par l'ETE, représente 9,1 % du vote du collège électoral (ce qui correspond à un peu plus de 2 % du vote de l'ensemble du corps électoral). L'erreur à l'intérieur de l'échantillon, c'est-à-dire l'erreur moyenne absolue de prédiction sur les onze élections, ne représente que 5,6 % du vote du collège électoral (ce qui correspond à un peu plus de 1 % du vote de l'ensemble du corps électoral). Ces basses estimations d'erreur nous ont encouragés à appliquer le modèle en dehors de l'échantillon, pour prévoir les résultats des élections proches de 1992. Les valeurs de 1992 pour les variables de prédiction étaient les suivantes : G = 1,1 ; P = 32 ; S = - 8 ; C = 1, ce qui donnait la prévision suivante, disponible au milieu de l'été 1992 :

$$\begin{aligned} V &= 6,83 + 7,76 (1,1) + 0,86 (32) + 0,53 (- 8) + 19,66 (1) \\ &= 58,39 \% \text{ du vote du collège électoral en faveur de Bush.} \end{aligned}$$

Autrement dit, le modèle prédit une légère victoire de Bush, semblable à celle de Truman en 1948, lorsqu'en tant que candidat sortant il reçut 57,1 % des voix du collège électoral. Manifestement, il n'en a pas été ainsi. Bush a reçu en fait 31 % des voix du collège électoral, ce qui équivaut à une erreur de + 27 %. Il est difficile de considérer cette erreur, égale à trois fois la taille du ETE (qui est de 9,1), comme étant le fruit du hasard. Que s'est-il donc passé ? Une explication possible est que le modèle a été la victime d'exploitations antérieures de données. Ceci est peu probable étant donné les résultats des tests de stabilité et autres diagnostics. Une explication immédiate est que la variable dépendante devrait être la mesure des voix de l'ensemble du corps électoral, mesure plus traditionnelle, et devrait se limiter à la différence entre les voix des deux partis (au lieu des voix du parti entier). Cependant, de telles expériences de mesure ne changent pas les résultats de manière significative. Une autre explication possible est l'erreur de spécification. Il est tout particulièrement possible que certaines variables, notamment la force du parti (S) et l'attrait du candidat (C) soient des facteurs inappropriés. (Il se peut que ces variables soient capturées d'une manière effective par les autres variables qui sont déjà dans le modèle.) Voici donc le modèle déjà présenté, mais sans les variables suspectes S et C, utilisé pour prédire les élections de 1992 :

$$\begin{aligned} \text{Equation 2} \quad V &= - 20,90 + 6,83* G + 1,40* P \\ &(- 1,26) (2,04) \quad (3,75) \end{aligned}$$

R-carré = 0,81

R-carré corr. = 0,76

N = 11

ETE = 15,52

D - W = 1,63

où les variables sont les mêmes que dans l'équation précédente.

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

Si l'on utilise ce dernier modèle pour prédire le résultat des élections de 1992, on obtient :

$$\begin{aligned} V &= - 20,90 + 6,83 (1,1) + 1,40 (32) \\ &= - 20,90 + 7,51 + 44,8 \\ &= 31,41 \% \text{ des voix du collège électoral en faveur de Bush} \\ &= 169 \text{ des voix du collège électoral en faveur de Bush.} \end{aligned}$$

### D) Évaluation de la précision du Modèle I

Ce modèle révisé, appelé Modèle I (ce qui aide à indiquer que c'est un sous-ensemble du Modèle II), a des caractéristiques attirantes. Dans ce cas précis, la prévision a différé du nombre véritable de voix du collège électoral de seulement une voix (Bush a en fait recueilli 168 des voix du collège électoral). Par ailleurs, cela suggère une théorie générale minutieuse. Il est supposé que l'électeur se livre à un calcul basé sur la manière dont l'administration sortante s'est occupée des problèmes économiques et non économiques. Les variables appropriées sont facilement mesurables, en matière de croissance économique et de popularité du Président. Enfin, le modèle I est convenablement dénommé, puisqu'il suit essentiellement notre formulation originale (Lewis-Beck et Rice, 1984 ; 1992, p. 47).

Dans la mesure où l'estimation du Modèle I est plus juste, il devrait obtenir de meilleurs résultats que ses rivaux. Dans le tableau 3, on trouve trois modèles rivaux. La première colonne présente une formulation a-théorique, avec le vote présenté comme étant un processus de séries chronologiques "de long décalage géométrique". L'idée est de gagner un maximum d'information statistique en examinant l'histoire des données, le résultat des dernières élections et la popularité du Président pendant les dix-huit derniers mois. La deuxième colonne contient un modèle qui allie la théorie du Modèle II à la notion de décalage géométrique. Bien que la théorie de ce modèle soit sans doute "faible", elle essaie de profiter de tous les degrés possibles de liberté. La troisième colonne présente le Modèle I, appelé "théorie forte". La quatrième colonne postule que l'élection du Président se fait en fonction de la satisfaction du public, comme il a été mesuré dans le sondage Gallup réalisé le mois précédant les élections. Ce modèle est "naïf", disant simplement que si l'on aime la manière dont le Président fait son travail, on vote pour lui ou le candidat de son parti.

Est-ce que les modèles sont comparables en termes de capacité prédictive des résultats ? Examinons d'abord les prévisions faites pour 1992 (voir l'avant-dernière ligne dans le tableau 3). La "théorie forte" de la colonne 3 a obtenu les meilleurs résultats, alors que le "long décalage géométrique" a-théorique a obtenu les pires. Mais si l'estimation du Modèle I (colonne 3) est vraiment plus proche de la vraie structure, alors elle devrait pouvoir régulièrement faire de meilleures prévisions. Pour vérifier cette possibilité, j'ai réévalué chacun des modèles, en utilisant seulement les élections ayant eu lieu entre 1948 et 1980. À partir de ces équations réévaluées (chaque N = 9), j'ai prédit le résultat des élections de 1984, 1988 et 1992. Le modèle de "forte théorie" a donné les meilleures prévisions, avec la plus petite erreur moyenne de prévision (voir la dernière ligne, tableau 3).

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

**Tableau 3.** Modèles rivaux de prévision des élections présidentielles, MCO,  
N = 11 (le nombre d'élections entre 1848 et 1988)

	Grand décalage géométrique	Théorie faible	Théorie forte	Naïf
Décalage de vote	- 0,09	0,05		
Pop t-1				1,84
Pop t-4	1,57	- 0,45	1,40	
Pop t-6	2,01	3,03		
Pop t-9	- 1,92	- 3,82		
Pop t-12	2,72	7,29		
Pop t-18	- 2,92	- 4,54		
Econ		11,19	6,83	
Partyx		1,18		
Candx		- 44,42		
Constante	- 8,43	13,21	- 20,90	- 33,10
R-carré corr.	0,68	0,99	0,76	0,49
ETE	18,35	3,77	15,82	22,93
Erreur de 1992 (31, 23-prévision)	- 71,83	- 60,59	- 0,18	- 1,77
Erreur médiane	- 23,87	85,88	- 20,15	- 24,29
<p><b>Note :</b> Le décalage de vote = le pourcentage de voix du collège électoral reçues par le président sortant aux dernières élections ; Pop = le pourcentage de personnes interrogées approuvant le Président dans un sondage Gallup, avec un décalage allant de 4 à 18 mois avant les élections ; Econ = le taux de croissance du PNB ; Partyx = le nombre précis de sièges gagnés ou perdus par le président sortant aux dernières élections législatives ; Candx = la force du candidat sortant. (Ces trois dernières variables se mesurent toutes comme dans l'équation 1.)</p>				

Le Modèle I donne de meilleurs résultats que les autres modèles à ce test difficile, dans lequel des prévisions faites avant les élections sont plus tard refaites. Il semblerait donc que ce modèle soit un instrument prometteur pour prédire les résultats de la course à la présidence de 1996. Ci-dessous, je mets à jour le Modèle I, en incluant l'information de 1992 à l'intérieur de la série, pour avoir un N plus grand (= 12) :

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

**Equation 3** 
$$V = -20,97 + 6,83 * G + 1,40 * P$$
$$(-1,44) (2,17) (4,13)$$

R-carré = 0,82  
R-carré corr. = 0,78  
N = 12  
ETE = 14,91  
D - W = 1,75

où les variables et les statistiques sont les mêmes que dans l'équation 2.

Notons la stabilité des coefficients, des écart-types et des statistiques fiables par rapport à l'équation 2. Cette stabilité est une autre preuve que ces variables, l'économie et la popularité, sont des facteurs appropriés dans un vrai modèle de prévision des élections présidentielles.

### 3. Prédire les élections en France

En France, il existe beaucoup de recherches relatives à la popularité du Président sur séries chronologiques (voir surtout les nombreux ouvrages de Jean-Dominique Lafay et Jean-Luc Parodi). Le lien général important entre l'économie et le résultat des élections est par ailleurs reconnu depuis un certain temps (Rosa et Amson, 1976). Plus précisément, les conditions économiques sont depuis longtemps liées aux prévisions électorales, du moins dans l'arène législative (Lafay, 1977). Cependant, en ce qui concerne les modèles de prévision des élections présidentielles, peu de recherche a été faite, sans doute à cause du problème posé par la petitesse de l'échantillon. Après tout, sous la V<sup>e</sup> République, il n'y a eu que cinq élections présidentielles. Tout de même, comme on le verra, on peut glaner beaucoup à partir de ces cinq élections. Mais avant de procéder à notre analyse, considérons d'abord la prévision des élections à l'Assemblée Nationale, où les cas sont nombreux.

#### A) Prévision des élections à l'Assemblée Nationale

Supposons l'existence d'un modèle de base, presque semblable à celui utilisé pour les élections aux États-Unis, pour prédire le résultat des élections françaises. Selon ce modèle, le support des candidats sortant peut être prédit à partir d'indicateurs de questions économiques et non économiques, mesurés avec une avance raisonnable avant les élections. En 1985, j'ai estimé (MCO) un tel modèle comme il suit (Lewis-Beck, 1985b) :

**Equation 4** 
$$V = 62,7 - 3,29 G - 0,24 P$$
$$(7,50) (-1,03) (-1,34)$$

R-carré = 0,71  
R-carré corr. = 0,57  
N = 7  
ETE = 3,18  
D - W = 2,32

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

avec  $V_t$  = le pourcentage de l'ensemble des votes reçus par tous les partis de l'opposition à la coalition du parti du Président lors du premier tour des élections législatives ;  $G$  = le taux de croissance réelle du PIB au cours du trimestre précédant de six mois les élections ;  $P$  = le pourcentage de personnes ayant exprimé leur satisfaction envers le Président dans un sondage IFOP effectué six mois avant les élections ; les chiffres entre parenthèses sont les écart-types ;  $R$ -carré = le coefficient de détermination multiple ; ETE = l'écart -type de l'estimation de  $V$  ;  $D - W$  = les statistiques de Durbin-Watson ;  $N$  = le nombre d'observations (7 pour les élections comprises entre 1958 et 1981).

La fiabilité du modèle,  $R$ -carré = 0,71, n'est pas loin du modèle de base américain de l'équation 2. L'erreur de prévision à l'intérieur de l'échantillon pour ces sept élections est d'une moyenne de seulement 1,76 %. Par ailleurs, le modèle a toujours prédit correctement la coalition obtenant la majorité (la droite *versus* la gauche) et ceci même dans le cas du mouvement vers la gauche en 1981, cas qui, à l'époque, était unique. Enfin, l'application du modèle a produit une prévision exacte, avant les élections, de la défaite de la gauche en 1986, défaite qui a donné naissance à la première période de cohabitation (Lewis-Beck, 1986). Ci-dessous, on trouve un modèle (estimation MCO) mis à jour pour inclure des observations sur les élections de 1986, ainsi que sur celles de 1988 :

**Equation 5**  $V = 66,83 - 0,73 G - 0,34 P$

$(9,26) \quad (-0,22) \quad (-2,02)$

$R$ -carré = 0,65  
 $R$ -carré corr. = 0,53  
 ETE = 3,65  
 $N = 9$   
 $D - W = 1,48$

où les statistiques et les variables sont définies comme dans l'équation 4.

La fiabilité des statistiques montre peu de changement avec la formulation précédente dans l'équation 4. On peut observer une certaine instabilité des coefficients, apparemment due à une grande collinéarité entre la croissance économique et la popularité du Président (entre ces deux variables,  $r = 0,75$ ). En raison de cette collinéarité, il est difficile de distinguer les effets individuels de l'économie et de la popularité. (En elle-même, l'économie a une corrélation de  $-0,64$  avec les élections ; en elle-même, la popularité a une corrélation de  $-0,80$  avec les élections.) Cependant, l'effet collectif des variables, considérées ensemble, peut être évalué et des prévisions peuvent être faites. Quels sont les résultats du modèle, basé sur les élections entre 1958 et 1988, en ce qui concerne une prévision effectuée avant les élections de 1993 ? Ci-dessous, nous entrons les valeurs appropriées de 1993 pour les variables de prédiction :  $G = 1,2$  ;  $P = 31$ .

$$\begin{aligned} V(93) &= 66,83 - 0,73 G - 0,34 P \\ &= 66,83 - 0,73 (1,2) - 0,34 (31) \\ &= 66,83 - 0,88 - 10,54 \\ &= 55,4 \% \text{ des voix pour les partis de l'opposition de l'époque.} \end{aligned}$$

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

En d'autres termes, le modèle prédit une victoire certaine de la droite. Les partis de droite (RPR, UDF, FN et Divers droite) ont en fait reçu 57 % des voix au premier tour des élections à l'Assemblée Nationale en 1993. Par conséquent, le modèle a une erreur de prévision *ex ante* de seulement 1,6 % pour cette élection. (Notez que cette erreur est proche de la moyenne à l'intérieur de l'échantillon précédemment rapportée.) En fait, étant donné l'étendue inattendue de la défaite des socialistes en 1993, la précision du modèle est plutôt surprenante. Ci-dessous, je révisé le modèle une nouvelle fois, en y incorporant les observations de 1993.

**Equation 6**                       $V = 68,28 - 0,13 G - 0,38 P$   
(14,73) (-0,05) (-3,69)

R-carré = 0,73

R-carré corr. = 0,65

ETE = 3,40

N = 10

D - W = 1,82

où les variables et les statistiques sont définies comme dans l'équation 5.

Le modèle, comme il a été estimé (MCO) dans l'équation 6, utilise des observations faites sur toute la série des élections à l'Assemblée Nationale, ayant eu lieu sous la V<sup>e</sup> République, depuis 1958 jusqu'à 1993. Il est clair que c'est un outil de prévision prometteur pour les prochaines élections législatives.

### B) Prévision des élections présidentielles françaises

L'incursion que nous venons de faire dans les prévisions des élections législatives suggère qu'une concentration sur le modèle de base économie-popularité est profitable dans le contexte français, même avec un petit nombre de cas. En ce qui concerne les élections présidentielles françaises, il est indéniable que le nombre de cas – cinq – est petit (1965, 1968, 1974, 1981 et 1988). Cependant, si la théorie est forte et les données sont bonnes, on peut accomplir beaucoup. Les résultats initiaux sont certainement encourageants. Dans le tableau 4, on trouve de simples corrélations entre les mesures de popularité, d'économie et de vote aux élections présidentielles. [Pour plus de détails, voir Lewis-Beck et Rice (1992, p. 123).]

Plusieurs choses méritent d'être notées. Premièrement, la popularité du Président a un grand rapport avec les voix. Deuxièmement, la croissance économique, aussi, a un grand rapport avec les voix. Troisièmement, les mesures avec l'avance la plus grande pour les prévisions, c'est-à-dire les mesures effectuées environ six mois avant les élections, montrent de grandes corrélations avec le vote effectif. En particulier, la popularité (t - 6) corrèle à 0,77 et l'économie (t - 6) corrèle à 0,79. Quatrièmement, ces deux corrélations sont très proches des corrélations comparables de l'exemple des États-Unis. En effet, pour les États-Unis, le résultat des élections (V) en corrélation avec la popularité (P) et la croissance économique (G), produit respectivement  $r = 0,85$  et  $r = 0,69$  (toutes les variables sont mesurées avec l'équation 3).

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

**Tableau 4.** Corrélations (r) entre différentes mesures de la popularité du Président sortant, de la croissance économique et du vote pour le Président au second tour, pour les élections ayant eu lieu sous la V<sup>e</sup> République (de 1965 à 1988)

	<b>Vote</b>	<b>Pop (t)</b>	<b>Pop (t-6)</b>	<b>Econ (T)</b>	<b>Econ (T-1)</b>
<b>Pop (t)</b>	0,68				
<b>Pop (t-6)</b>	0,77	0,96			
<b>Econ (T)</b>	0,95	0,82	0,84		
<b>Econ (T-1)</b>	0,57	0,88	0,92	0,66	
<b>Econ (t-6)</b>	0,79	0,92	0,98	0,85	0,95

**Note :** Vote = le pourcentage de voix reçues au second tour par le président sortant (ou le candidat de son parti ou de sa coalition) ; Pop = le pourcentage de personnes "satisfaites" du Président selon un sondage effectué par l'Institut Français d'Opinion Publique (IFOP) (t = le sondage le plus proche de t - 6 = le sondage effectué six mois avant les élections) ; Econ = le pourcentage de taux de croissance du vrai PIB (T = année des élections ; T - 1 = l'année précédant les élections ; t - = le trimestre six mois avant les élections).

Supposons l'existence d'un modèle de base de prévision pour les élections présidentielles françaises presque semblable à celui des États-Unis :

Vote = les conditions économiques + la popularité du Président sortant

Ce modèle est estimé (MCO) dans le tableau 5. Alors que la fiabilité du modèle n'est pas mauvaise ( $R$ -carré = 0,62), aucun des coefficients n'offre une grande signification statistique, en raison d'un grand problème de collinéarité. (La corrélation entre l'économie et la popularité est de 0,98.) Par conséquent, en même temps que ce modèle multivarié, les modèles bivariés de ces variables sont représentés.

En général, quel est le taux de réussite de prévision de ces modèles ? La moyenne absolue d'erreur à l'intérieur de l'échantillon (EMA, tableau 5) est relativement basse, à 1,4 % ou 1,5 %. Est-ce que cette erreur est suffisamment petite pour prévoir des changements politiques majeurs ? En particulier, est-ce qu'elle peut prédire si le candidat sortant restera ou non au pouvoir ? Oui, quelle que soit l'estimation, qu'elle soit bivariée ou multivariée, la prévision annonce toujours le sort du président sortant. Ce qui est particulièrement significatif, c'est que chacun de ces modèles rapporte le triomphe sans précédent de la gauche en 1981. Somme toute, il semble possible de prédire le résultat des élections présidentielles françaises avec une certaine exactitude.

Quel est le pronostic pour les élections présidentielles de 1995 ? Supposons l'exactitude des données concernant la popularité et l'économie, et disponibles en

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

**Tableau 5.** Modèles de prévision pour les élections présidentielles françaises  
(le pourcentage de voix reçues par le candidat sortant au second tour des élections)

Facteur de prédiction	(1)	(2)	(3)
Econ (t-6)	3,86 (2,23)		3,99 (0,37)
Pop (t-6)		0,47 (2,11)	- 0,02 (- 0,01)
Constante	48,02 (19,24)	27,44 (2,26)	48,77 (0,82)
R-carré	0,62	0,60	0,62
R-carré corr.	0,50	0,46	0,25
ETE	2,56	2,65	3,14
EMA	1,47	1,41	1,47
D-W	3,24	3,04	3,24
N	5	5	5
Prévision pour 1995	52,0	48,6	51,9

**Note :** Les variables sont mesurées de la même manière que dans le tableau 4. Les statistiques sont définies de la même manière que dans les équations données plus haut, toutes estimées par les MCO. EMA = l'erreur moyenne de prédiction absolue à l'intérieur de l'échantillon. Les prévisions pour 1995 sont basées sur Pop (t - 6) = 45 (le pourcentage de personnes interrogées ayant répondu qu'elles étaient "satisfaites" du Président François Mitterrand dans le sondage IFOP réalisé en octobre 1994) et sur Econ (t - 6) = 1,02 (le pourcentage de changement du PIB, non annuel, pendant le troisième trimestre de 1994. Plus précisément, la moyenne des trois mois disponible comparée à la moyenne des trois derniers mois, publiée dans le numéro de la semaine du 15 au 21 octobre de *The Economist*, p. 138. C'est donc une première estimation du troisième trimestre).

octobre 1994, popularité (t - 6) = 45, économie (t - 6) = 1,02 (voir la note du tableau 5 pour plus de détails). Ces chiffres sont assez bons en ce qui concerne la manière dont ils sont comparables avec les autres chiffres de la série. Par exemple, ils sont pires que ceux avant les élections de 1988, lorsque le Président sortant François Mitterrand a remporté les élections (pour 1988, les chiffres respectifs étaient 56 et 1,5). Cependant, les chiffres sont meilleurs qu'avant les élections de 1981, quand le Président sortant Giscard d'Estaing a perdu (pour 1981, les chiffres respectifs étaient 44 et



## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

0,00). Lorsque les valeurs courantes pour l'économie et la popularité sont entrées dans les modèles, les prévisions faites avant les élections de 1995 sont effectuées (comme il l'est rapporté dans le tableau 5). Chaque modèle prédit une course serrée avec le candidat socialiste recevant entre 48,6 et 52 % des voix au second tour.

Laquelle de ces prévisions est la plus susceptible d'avoir raison ? Des trois modèles, je préfère le multivarié. Pour une première raison : les modèles bivariés représentent une commission délibérée d'erreur de spécification. Pour une autre raison : la collinéarité élevée du modèle multivarié, alors qu'elle crée un problème d'instabilité du coefficient, ne produit pas en lui-même des calculs au paramètre biaisé. En fin de compte, le modèle économie-popularité déterminé plus convenablement devrait donner de meilleures prévisions. (Par ailleurs, comme des observations supplémentaires sont ajoutées à la série, le problème de collinéarité devrait diminuer.) Ce modèle prédit une victoire socialiste avec 51,9 % des voix.

Quelles sont les chances que cette prévision soit juste ? Puisque, pour toute prévision électorale, quelque erreur est pratiquement inévitable, il est presque certain qu'elle ne sera pas exactement juste. Jusqu'à quel point se trompera-t-elle ? L'erreur standard d'estimation (3,14), appliquée rigoureusement, suggère qu'elle pourrait se tromper de beaucoup. Mais sur toute l'histoire des élections présidentielles de la V<sup>e</sup> République, l'erreur moyenne de prévision (EMA) n'est que d'environ 1,5 %. En prenant ce chiffre comme grossière règle générale, nous ne devrions pas être surpris si les résultats de 1995 tombaient jusqu'à 50,4 %. Ceci permettrait toujours une victoire socialiste avec une marge des plus faibles.

### 4. Conclusion

Les modèles ont une valeur considérable en tant que méthode de prévision électorale, particulièrement quand on les compare aux intentions de vote pré-électorales des sondages. En se servant d'une théorie de conduite démocratique bien fondée, des modèles de base très forts du point de vue statistique pour des prévisions d'élections présidentielles, peuvent être développés. Il semblerait que ces modèles théoriquement simples et empiriquement minutieux pourraient être appliqués à de nombreuses démocraties occidentales industrialisées. Les tests particuliers offerts au sujet de cultures différentes traitent de la France et des États-Unis. Dans chacun de ces pays, il est souvent possible de prédire le vainqueur des élections présidentielles plusieurs mois à l'avance, en utilisant des mesures comparables de croissance économique et de popularité du Président.

Est-ce que cela veut dire que les modèles donnent des prévisions parfaites ? Absolument pas. Ces modèles représentent un bon point de départ, mais il y a toujours erreur. Cette erreur peut être due, entre autres choses, à l'exclusion de variables importantes ou au hasard. Notre tâche est de réduire l'erreur qui est facile à trouver, la partie systématique. La difficulté indicative, à mon avis, est toute erreur d'estimation restante alliée à un problème de petitesse de l'échantillon. Quand des variables

## COMPARAISON DE PRÉVISION DES ÉLECTIONS PRÉSIDENTIELLES

supplémentaires sont ajoutées au modèle de base (économie-popularité), seule la "théorie forte" devrait justifier cette inclusion. Puisque peu de variables peuvent passer avec succès le test de "théorie forte", peu seront ajoutées. Cependant, ceci est pour le mieux, puisque de petits échantillons demandent que les additions soient limitées, de manière à ne pas augmenter la "chance" de fiabilité du modèle. Par ailleurs, le maintien d'une nouvelle variable dans le modèle finirait par demander une preuve que, sur des essais répétés (les élections), elle aide vraiment à réduire l'erreur de prévision. Ainsi, les ajustements d'estimation à partir du modèle de base continueraient avec prudence, à la manière longue.

### BIBLIOGRAPHIE

- BRODY Richard, SIGELMAN Lee, "Presidential Popularity and Presidential Elections: An update and an extension", *Public Opinion Quarterly*, 47, 1983, pp. 325-328.
- CAMPBELL James E., MANN Thomas E., "Forecasting the 1992 Presidential Election", *The Brookings Review*, 1, 1992, pp. 22-27.
- LAFAY Jean-Dominique, "Les conséquences électorales de la conjoncture économique : essais de prévision chiffrée pour mars 1978", *Vie et Sciences Économiques*, 75, 1977, pp. 1-7.
- LEWIS-BECK Michael S., "Election Forecasts in 1984: How accurate were they?", *PS*, 18, 1985a, pp. 53-62.
- LEWIS-BECK Michael S., "Un modèle de prévision des élections législatives françaises (avec une application pour 1986)", *Revue Française de Science Politique*, 35, 1985b, pp. 1080-1091.
- LEWIS-BECK Michael S., "Les législatives de 1986 : Nouveau clivage ou restauration ?", *Politique Économique*, 3, 1986, pp. 31-33.
- LEWIS-BECK Michael S., RICE Tom W., "Forecasting Presidential Elections: a Comparison of Naive Models", *Political Behavior*, 6, 1984, pp. 9-21.
- LEWIS-BECK Michael S., RICE Tom W., *Forecasting Elections*, Washington, D.C., CQ Press, 1992.
- POLSBY Nelson W., WILDAVSKY (Aaron), *Presidential Elections: Strategies of American Electoral Politics*, New York, Scribner, 1984.
- ROSA Jean-Jacques, AMSON Daniel, "Conditions économiques et élections : une analyse politico-économétrique (1920-1973)", *Revue Française de Science Politique*, 26, 1976, pp. 1101-1124.
- TUFTE Edward R., *Political Control of the Economy*, Princeton, Princeton University Press, 1978.