

Modèles de prévision des élections au Québec

Intelligence d'affaires

**Mémoire en vue de l'obtention
du grade de Maitrise en science (Msc)**

Décembre 2013

© Amina Chiadmi, 2013

Sommaire

Le présent mémoire porte sur l'élaboration et la comparaison de plusieurs modèles statistiques pour la prévision des élections québécoises. Nos modèles ont été basés sur plusieurs recherches statistiques qui ont permis d'étudier les différents modèles réalisés par les chercheurs et de déduire les facteurs qui peuvent influencer le résultat des élections.

Le premier modèle tente de prévoir le nombre de votes en ajustant les résultats des dernières élections avec les sondages effectués quelques semaines avant les élections. Le deuxième modèle a pour objectif de transformer les intentions de vote en pourcentage de vote. Le troisième modèle établit une prévision du pourcentage de vote pour chaque circonscription du Québec alors que le quatrième modèle prévoit le taux d'erreur que les sondages peuvent produire. Ce taux d'erreur ajuste les résultats de sondages et permet de prévoir convenablement les résultats des élections.

Malgré que la variable à expliquer diffère d'un modèle à un autre, tous les modèles ont comme résultat final l'estimation du nombre de sièges que chaque parti politique pourra détenir lors des prochaines élections. Le modèle le plus performant, qui sera le plus proche des résultats réels, sera choisi pour des prévisions futures.

Remerciements

Je tiens à remercier mes deux directeurs de recherche, Marc Fredette et Denis Larocque, pour leur implication dans la réalisation de ce mémoire. Ils étaient présents et interactifs tout au long de la réalisation du projet de recherche, de l'idée de départ jusqu'aux corrections finales.

Je remercie également l'entreprise Awane Insight, qui m'a accueilli pour réaliser ce projet ainsi que Mitacs qui a financé en partie mon stage au sein de l'entreprise.

Je remercie ma famille, qui m'a soutenu tout au long de la réalisation du mémoire. Je dédie ce mémoire à mon mari, à mes deux enfants qui ont vécu avec moi cette expérience avant même qu'ils ne voient le jour.

Table des matières

Sommaire.....	i
Remerciements.....	ii
Table des matières.....	iii
Liste des tableaux.....	v
Introduction	1
Chapitre 1: Revue de littérature	3
1.1 Aperçu sur les élections québécoises	3
1.1.1 Généralités.....	3
1.1.2 Le système électoral au Québec :	4
1.1.3 Les circonscriptions et découpage électoral.....	4
1.1.4 Les sondages et intentions de vote.....	5
1.1.4.1 Incidences directes :	5
1.1.4.2 Incidences indirectes :	5
1.1.5 Les institutions de sondages :	6
1.1.6 Aperçu sur les élections de 2007, 2008 et 2012	6
1.2 Les modèles prévisionnels des élections	8
1.2.1 Vue d'ensemble	8
1.2.2 Modèle Rosenthal	10
1.2.2.1 Présentation du modèle :	10
1.2.2.2 Description de la fonction.....	11
1.2.2.3 Principaux résultats.....	12
1.2.2.4 Application au niveau du Québec :	13
1.3 Les facteurs qui influencent le vote	14
1.3.1 Les facteurs sociaux démographiques:	14
1.3.2 Les facteurs à court terme	15
1.3.2.1 Les enjeux.....	15
1.3.2.2 L'image des chefs :	16
1.3.2.3 La conjoncture économique :	16
Chapitre 2 : Méthodologie de recherche.....	18
2.1 Les données de l'étude	18

2.1.1	Les données utilisées:	18
2.1.2	Description des variables:	20
2.1.3	Préparation des données :	20
2.2	Description des modèles de prevision:	20
2.2.1	Le modèle de base de Rosenthal :	22
2.2.2	Le modèle basé uniquement sur les sondages :	22
2.2.3	Le modèle de prévision du pourcentage de vote	23
2.2.4	Modèle de prévision de l'erreur de sondage :	24
2.3	Principes de base de la prévision :	25
2.4	Démarches statistiques	26
2.4.1	La Régression linéaire multiple:	26
2.4.2	La sélection de variables:	27
a.	La procédure GLMSELECT :	27
b.	Méthodes de sélection de variables :	28
c.	Les critères de sélection des variables :	29
2.4.3	Comparaison des prévisions avec le résultat réel :	30
2.4.4	Les indicateurs de performance :	30
Chapitre 3 : Analyses et résultats		32
3.1	Résultat du modèle de base de Rosenthal :	32
3.2	Résultat du modèle basé uniquement sur les sondages :	33
3.3	Résultat du modèle de prévision du pourcentage de vote:	34
3.3.1	Les variables significatives	34
3.3.2	Prévision du nombre de sièges :	34
3.4	Résultat du modèle de prévision de l'erreur de sondage:	35
3.4.1	Variables significatives	35
3.4.2	Prévision du nombre de sièges	35
3.5	Tableau récapitulatif des résultats des 4 modèles	36
3.6	Analyse des résultats des modèles de prévision :	37
Conclusion :		39
Références bibliographiques :		41
Annexes		42

Liste des tableaux

Tableau 1 : pourcentage de vote et résultats du dernier sondage avant les élections provinciales de 2007, 2008 et 2012	7
Tableau 2: modèle de base de Rosenthal : Nombre de sièges par parti politique selon le modèle 1 et 2 sans effet de performance (N) et avec effet de performance (Y) ainsi que le résultat réel des élections 2011.....	12
Tableau 3: Résultat du modèle de base de Rosenthal pour la prévision des élections provinciales de 2008.	32
Tableau 4: Résultat du modèle de base de Rosenthal pour la prévision des élections provinciales de 2012.	32
Tableau 5: Taux de bonne classification du modèle de base de Rosenthal	33
Tableau 6: Nombre de sièges bien prédit pour le modèle basé uniquement sur les sondages	33
Tableau 7: Modèle de prévision du pourcentage de vote: Prévision du nombre de sièges par parti politique pour les élections de 2008.....	34
Tableau 8: Modèle de prévision du pourcentage de vote: Prévision du nombre de sièges par parti politique pour les élections de 2012	34
Tableau 9: Modèle de prévision de l'erreur de sondages: Prévision du nombre de sièges par parti (2008)	35
Tableau 10: Modèle de prévision de l'erreur de sondages: Prévision du nombre de sièges par parti (2012)	36
Tableau 11: Tableau récapitulatif des résultats des 4 modèles pour les élections de 2008..	36
Tableau 12: Tableau récapitulatif des résultats des 4 modèles pour les élections de 2012..	37

Introduction

Prévoir quelques semaines d'avance le parti politique qui va gagner les élections n'est pas quelque chose de nouveau, ni de dépassé. Les firmes de sondages utilisent les techniques de sondages afin de mesurer les intentions de vote des citoyens. Des chercheurs utilisent des méthodes statistiques plus poussées afin de bâtir des modèles de prévisions. La majorité de ces modèles ont été élaborés afin de prévoir les résultats des élections présidentielles aux États-Unis et peu de recherches ont été réalisées dans ce sens au niveau du Canada en général, et du Québec en particulier.

Le mémoire a pour but de construire et comparer des modèles de prévisions du nombre de sièges que va détenir chaque parti politique lors des prochaines élections.

Le présent mémoire a été réalisé dans le contexte d'un projet en entreprise. Les données ainsi que les résultats obtenus vont être utilisés également pour bâtir un profil pour chaque circonscription du Québec. Un projet qui permet aux partis politiques d'orienter leurs choix stratégiques en matière de campagne électorale.

Dans ce présent mémoire, nous allons mettre l'accent sur le développement des modèles de prévisions. Quatre modèles ont été développés en se basant sur les données des élections de 2007, 2008 et 2012. Le modèle le plus performant sera choisi comme modèle de prévisions des élections futures.

L'étude des différents modèles qui ont été réalisés dans le passé est la première étape de notre recherche. Cela a permis non seulement de mettre en revue les différentes méthodes ou fonctions statistiques utilisées par les chercheurs mais aussi de déduire les facteurs qui peuvent avoir un impact sur les résultats des élections ainsi que les facteurs qui peuvent influencer le vote des électeurs.

Suite à cette étude, nous avons construit les variables et organisé les données collectées. Ces variables ont été utilisées totalement ou partiellement selon le modèle élaboré. Nous allons également présenter dans la méthodologie de recherche la modélisation avec la régression linéaire, méthode utilisée afin de construire deux des quatre modèles de prévisions.

Le choix du meilleur modèle de prévision des élections québécoises est basé sur l'étude et la comparaison de plusieurs critères de performance que nous allons présenter dans la troisième et dernière partie du présent mémoire.

Enfin, nous concluons ce mémoire en procédant à un bref retour sur les quatre modèles de prévisions qui ont été construits, en mettant l'accent sur le modèle le plus performant. Nous allons également présenter les limites de recherches et les solutions et suggestions potentielles afin d'améliorer la performance du modèle choisi pour de futures études. Nous allons également suggérer d'autres pistes de recherche que l'entreprise pourrait prendre en considération pour exploiter le mieux possible les résultats de cette recherche.

Chapitre 1: Revue de littérature

La présente revue de littérature a comme objectif de mettre en contexte notre étude sur la modélisation des élections. Tout d’abord, nous allons effectuer un tour d’horizon sur le fonctionnement des élections québécoises. L’accent sera mis sur les élections de 2007, 2008 et 2012, objet de notre recherche. Par la suite, nous allons présenter quelques travaux reliés à la prévision des élections, en occurrence les modèles américains et canadiens. Nous allons exposer les objectifs de prévisions de chacun des travaux, les techniques utilisées ainsi que les facteurs pris en considération afin de réaliser ces modèles. Les différentes recherches que nous allons présenter se basent en majorité sur l’historique des élections, les sondages et le facteur économique. Nous allons donc étudier d’autres facteurs qui peuvent influencer le vote en se basant sur les différentes études menées au Québec.

1.1 Aperçu sur les élections québécoises

1.1.1 Généralités

Les élections générales au Québec sont tenues au maximum tous les cinq ans afin d’élire les membres de L’Assemblée Nationale. Ces membres sont les députés élus dans chacune des 125 circonscriptions issus de différents partis politiques.

Au Québec, plusieurs partis entrent dans la scène politique : des partis « traditionnels » tel que le Parti Libéral, les conservateurs (1867-1935), le Parti Québécois et des « tiers partis » tel que L’Action Démocratique du Québec (ADQ), Québec Solidaire et le Parti Vert. Le gouvernement est formé par le parti qui détient le plus de sièges à l’Assemblée Nationale. Les autres partis forment l’opposition. Depuis les élections de 2012, le Parti Québécois, sous la direction de Pauline Marois, forme un gouvernement minoritaire suite à la défaite du Parti libéral qui avait dirigé le Québec depuis 2003.

1.1.2 Le système électoral au Québec

Le système électoral ou le mode de scrutin est un processus qui permet d'élire les représentants dans le cadre des élections. Il existe plusieurs modes de scrutin. Le Québec ainsi que le reste du Canada opte pour le scrutin uninominal majoritaire à un tour : l'électeur vote pour un candidat d'un parti politique, qui se présente dans une circonscription. Le candidat qui reçoit le plus de vote dans une circonscription devient député et remporte un siège à L'Assemblée Nationale. Le parti politique qui recueille le plus de sièges remporte les élections.

Cependant, ce mode peut faire élire un parti politique qui peut se révéler minoritaire, comme ce fut le cas lors des élections québécoises en 2012. Ce type de gouvernement doit obtenir un soutien majoritaire, en s'entendant avec des partis différents pour chaque projet de loi ou vote de confiance. La durée de vie d'un gouvernement minoritaire est généralement plus courte que celle d'un gouvernement majoritaire. S'ils sont insatisfaits du gouvernement et sont prêts à aller en élection, les partis d'opposition peuvent le renverser et déclencher des élections.

1.1.3 Les circonscriptions et découpage électoral

Actuellement, la carte électorale du Québec délimite le contour de 125 circonscriptions. La carte est revue chaque deux élections pour tenir compte de l'évolution de la population et d'assurer la plus grande homogénéité socio-économique possible. L'annexe 1a présente la liste des 125 circonscriptions depuis le découpage électoral de 2001. En 2011, une nouvelle carte électorale est adoptée, des circonscriptions ont été ajoutées, supprimées ou renommées (voir annexe 1b). Comme nous le verrons au cours de ce mémoire, ce changement de délimitation aura une incidence significative sur la construction de nos modèles.

1.1.4 Les sondages et intentions de vote

D'une manière générale, les sondages d'opinion jouent un rôle important dans les élections et sont devenus indispensable lors des campagnes électorales. Selon La Direction de la Recherche Parlementaire (1994), les sondages pré-électorales ont une incidence directe et indirecte sur le processus électoral :

1.1.4.1 Incidences directes

Certains observateurs politiques estiment que les sondages influencent le comportement et le choix des électeurs et donc les résultats des élections. Les électeurs auront tendance à suivre l'opinion générale et à voter en faveur du candidat qui a le plus de popularité (vague populaire). Cet « effet d'entraînement » s'explique par l'envie des électeurs d'être du côté du vainqueur. Dans son article « Do Opinion Polls Influence Voters? » Patrick Martin (1987) affirme que les sondages stimulent l'effet d'entraînement et favorise le « vote stratégique ». Ainsi, certains électeurs vont voter pour leur deuxième choix en voyant que leur premier choix n'a pas de chance de gagner.

Les sondages peuvent motiver ou démotiver les gens à voter. Lorsque les électeurs sont certains que leur candidat ou parti favori va l'emporter, ils ne prennent pas la peine d'aller voter. D'autres électeurs iront voter pour prouver que les sondages ont tort. Les gouvernements et les partis politiques se servent également des sondages pour le choix de la date des élections.

1.1.4.2 Incidences indirectes

Selon La Direction de la Recherche Parlementaire (1994), les sondages, renforcés par la couverture médiatique, mettent l'accent sur l'image des chefs et détournent le débat des questions essentielles. Ainsi, on fait passer ces sondages pour des mesures de succès de chacune des campagnes électorales.

Les sondages peuvent également avoir une incidence positive ou négative sur la campagne électorale : ils peuvent influencer l'orientation médiatique ainsi que la motivation des donateurs et bénévoles.

1.1.5 Les institutions de sondages

Les sondages sont généralement reconnus par leur côté scientifique : *"Le sondage d'opinion semble relever à la fois de la science, forme moderne dominante de la vérité, et de l'opinion publique, forme moderne dominante de la légitimité."* (Martin, 1993).

La technique la plus utilisée pour effectuer des sondages est celle de l'échantillonnage et l'analyse statistique. Les instituts de sondages choisissent un échantillon représentatif de la population et leur posent des questions sur leur intention de vote, leur avis sur les chefs des partis et les principaux enjeux. Les données de sondages sont pondérées à l'aide des données sociodémographiques (LegerMarketing, 2012).

Dans la présente étude, nous avons utilisé les publications de l'institution Leger Marketing puisqu'elle couvre un échantillon important (3000 à 3500 répondants) et présente des résultats par région.

1.1.6 Aperçu sur les élections de 2007, 2008 et 2012

Nous allons passer en revue les élections québécoises depuis 2007 afin de mettre en contexte nos modèles de prévisions. Le tableau suivant résume le résultat global des élections par parti politique pour les élections de 2007, 2008 et 2012 :

	Élections 2007			Élections 2008			Élections 2012		
	Sondages	Résultats	sièges	Sondages	Résultats	sièges	Sondages	Résultats	sièges
PLQ	36%	33,1%	48	46%	42,1%	66	28%	31,20%	50
PQ	29%	28,4%	36	34%	35,2%	51	33%	31,95%	54
ADQ/CAQ	25%	30,8%	41	12%	16,4%	7	27%	27,05%	19
Autre	10%	7,5%	0	8%	6%	1	12%	9,8%	2

Source : <http://www.electionsquebec.qc.ca/francais/provincial/resultats-electoraux/elections-generales.php>

Tableau 1 : pourcentage de vote et résultats du dernier sondage avant les élections provinciales de 2007, 2008 et 2012

Les élections de 2007

Les élections de 2007 ont connu une montée de l'Action Démocratique du Québec (ADQ). Dirigé par Mario Dumont, le parti prive le Parti Libéral du Québec (PLQ) de la majorité à l'assemblée nationale. Malgré l'insatisfaction envers son gouvernement durant son premier mandat (l'incapacité d'accorder des baisses d'impôts), le premier ministre Jean Charest a été réélu pour un deuxième mandat.

La montée de l'ADQ n'est pas expliquée par une forte mobilisation des nouveaux électeurs. Le taux de participation, malgré qu'il a atteint 71,23% n'a pas connu une variation significative¹. Cette montée a été produite au détriment des deux partis « traditionnels » en occurrence le Parti Libéral et le Parti Québécois. (Bélanger & Nadeau, 2009).

Les élections de 2008

Le déclenchement des élections une année après celles de 2007 avait pour objectif de regagner la majorité des sièges pour le Parti Libéral du Québec. Le contexte était favorable pour le PLQ puisqu'en 2008, le parti a regagné la popularité. La crise financière qui a éclaté en 2008 a offert un enjeu principal que le Parti Libéral du Québec défendra au cours de la campagne électorale. Le déclin de l'ADQ (passage de 41 à 7 sièges en 2008) a été expliqué par l'incapacité du parti à convaincre la

¹ <http://www.electionsquebec.qc.ca/francais/tableaux/historique-du-taux-de-participation.php>

population de ses compétences à assumer leur rôle politique (Bélanger & Nadeau, 2009). Les élections de 2008 n'ont pas connu une participation électorale semblable à celle de 2007, le taux de participation ne dépassait pas les 58%.

Les élections de 2012

La 40^{ème} élection générale québécoise a été tenue en septembre 2012. Une année avant la fin du mandat du premier ministre Jean Charest.

Les évènements qu'a connus le Québec quelques mois avant les élections, principalement la crise étudiante, ont probablement eu une incidence significative sur le résultat des élections. Par conséquent, c'est le Parti Québécois, sous la présidence de Pauline Marois, qui remporte les élections et forme un gouvernement minoritaire. Les élections ont connu une très forte participation qui a atteint 74,6%.

Un nouveau parti politique est entré en scène lors des élections de 2012 suite à une fusion avec l'Action Démocratique du Québec (ADQ). Fondée en 2011 par François Legault, la Coalition Avenir Québec (CAQ) est un parti politique de centre droit qui a comme priorité la santé, l'éducation et le développement économique.

1.2 Les modèles prévisionnels des élections

1.2.1 Vue d'ensemble

Les modèles statistiques ayant comme objectif de prédire le résultat des élections ont commencé à apparaître aux États-Unis dans les années 70.

Kramer (1971) a été le premier chercheur à estimer une fonction de vote pour les élections américaines en utilisant le modèle de vote prospectif² : Son modèle est basé sur les performances antérieures des partis politiques ainsi que les conditions

² Modèle de vote prospectif: un modèle où les électeurs ne sont plus tournés vers le passé mais vers le futur. Si des conditions économiques satisfaisantes sont anticipées pour la période post-électorale, les électeurs reconduisent le parti sortant même s'il a obtenu de mauvaises performances économiques avant les élections.

économique du pays. Fair (1978) généralise le modèle de Kramer (1971) et développe un modèle théorique de comportement électoral pour analyser l'influence de la conjoncture économique sur les résultats des élections.

D'autres chercheurs ont intégré le facteur économique dans leur modèle statistique tel que Lewis-Beck et Rice (1984), Abramowitz (1988), Alesina, Londregan et Rosenthal (1996). Les variables économiques ont été évaluées différemment selon les chercheurs : Abramowitz (1988) a utilisé les données de la croissance du PIB deux trimestres avant la date des élections et les a annualisé en 2004. Alesina et al. (1996) ont utilisé les données relatives à la croissance du PIB une année avant les élections. Fair (1978) a utilisé la croissance du PIB des trois derniers trimestres qui précèdent la date des élections.

Le facteur politique a été également pris en considération dans le développement des modèles de prévision des élections. La popularité du président ou le parti sortant est la variable la plus utilisée (Brody et Sigelman, 1983, Lewis-Beck et Rice, 1984, Abramowitz, 1988). La popularité est mesurée grâce aux sondages pré-électoraux précédant les élections.

D'autres facteurs politiques ont été intégrés dans les modèles de prévisions et peuvent, selon les chercheurs, influencer le choix des électeurs américains. Hibbs (2000) et Fair (2002) prennent en considération les guerres telles que la guerre du Vietnam et la guerre de Corée. Lewis Beck et Tien (2000) incluent dans leur modèle l'indice de paix et prospérité.

Abramowitz (1988) a développé le « Time for change model ». L'hypothèse stipule que si les électeurs veulent le changement, ils ne vont probablement pas voter pour le président sortant. Si la cote du président est élevée et l'économie se porte bien, le président sortant va être réélu. Abramowitz a appliqué ce modèle pour les élections de 2000, 2004 et 2008 et a utilisé la popularité du président sortant, la croissance du PIB réel et la variation du taux de chômage pour bâtir son modèle.

Au Canada, Rosenthal (2011) examine si des modèles statistiques simples auraient pu prévoir la majorité conservatrice avant les élections fédérales de 2011 au Canada. Selon son article 'Was the conservative majority predictable?', plusieurs organisations ont tenté d'utiliser les résultats de sondages préélectoraux pour prédire le nombre de siège pour chaque parti mais peu d'entre elles ont pu prévoir la majorité conservatrice. Même les compagnies de sondage n'ont pas pu prédire cette majorité selon Rosenthal (2011). Il conclut qu'en utilisant adéquatement les données de l'élection précédente, la majorité conservatrice aurait pu être anticipée. L'auteur présente une méthode simple de projection de sièges pour les élections canadiennes. Ce modèle simple fournit des projections de sièges qui sont assez proches des résultats des élections 2011.

En s'inspirant de la méthodologie de Rosenthal (2011) nous avons appliqué son modèle de base au niveau provincial. L'approche de Rosenthal est présentée plus en détail dans la sous-section suivante.

1.2.2 Le modèle de base de Rosenthal

1.2.2.1 Présentation du modèle

L'article de Rosenthal (2011) examine si les modèles statistiques simples auraient pu prévoir la majorité conservatrice avant les élections fédérales de 2011 au Canada. Rosenthal conclut qu'en utilisant adéquatement les données de l'élection précédente, la majorité conservatrice aurait pu être anticipée. L'article présente une méthode de recherche simple de projection de sièges pour les élections canadiennes, constituée par les données de chaque circonscription et les résultats des élections précédentes. Il modifie ensuite ces projections en se basant sur les derniers sondages pré-électoraux.

L'auteur a commencé avec une fonction de base qui prend en considération les résultats des élections précédentes par circonscription pour chaque parti politique et les a ajusté avec les derniers sondages avant les élections.

1.2.2.2 Description de la fonction

Pour chacune des 308 circonscriptions et pour chaque parti politique, Rosenthal applique la fonction de vote suivante :

Modèle 1 :
$$V = w + (p - r)T$$

Modèle 2 :
$$V = wp/r$$

Rosenthal préfère le modèle (2) au modèle (1), mais prend en considération les deux modèles. Selon lui, les résultats des deux modèles seront généralement similaires.

- **Variable à expliquer :**

V= Nombre de votes du parti politique dans une circonscription

- **Variables explicatives :**

T= total des votes dans la circonscription lors de la dernière élection.

w = nombre de votes reçus par un parti donné lors de la dernière élection.

p= résultat de sondages : intention de vote pour le parti politique quelques semaines avant les élections à prévoir.

r= pourcentage de vote pour le parti politique lors de la dernière élection.

Selon le modèle, le parti politique qui détient le nombre de votes le plus élevé remporte le siège dans une circonscription. Rosenthal totalise le nombre de sièges prévu pour chaque parti pour toutes les circonscriptions du Canada et obtient par la suite un résultat global.

Cette fonction de base a été améliorée par l'ajout de la « performance » globale de chaque parti politique. Certains partis sont bien connus pour être mieux que

d'autres à inciter leurs supporters à se rendre au bureau de vote le jour du scrutin. Leurs électeurs sont connus pour être plus «engagés» et donc plus susceptibles d'aller voter (et non pas changer d'avis). Ces effets peuvent conduire certains partis à surperformer par rapport à d'autres partis.

Pour tenir compte de ce facteur, Rosenthal a ajusté la fonction de base en estimant la tendance de chaque parti à surperformer (ou sous-performer). Il utilise les élections précédentes (2008) comme guide et calcule la différence entre les résultats des élections de 2008 et les résultats de sondages pré-électorales afin d'obtenir un pourcentage de sur performance ou sous performance (si le résultat est négatif) correspondante.

1.2.2.3 Principaux résultats

Le tableau suivant (Rosenthal 2011) reprend les résultats globaux par modèle et les compare avec le résultat réel des élections :

	Modèle 1 (N)	Modèle 1 (Y)	Modèle 2 (N)	Modèle 2 (Y)	Résultat réel
Conservateurs	147	157	148	160	166
Libéraux	47	40	37	33	34
NPD	109	104	111	103	103
Bloc québécois	5	7	12	12	4
Parti vert	0	0	0	0	1

Tableau 2: modèle de base de Rosenthal : Nombre de sièges par parti politique selon le modèle 1 et 2 sans effet de performance (N) et avec effet de performance (Y) ainsi que le résultat réel des élections 2011.

A partir de ce tableau de résultats, les résultats des modèles (1) et (2) sont similaires pour les conservateurs, cependant, les différences semblent plus marquées pour les autres partis. Si l'effet de performance n'est pas inclus, le nombre de sièges des conservateurs prévu est de 147 ou 148, un nombre qui n'atteint pas la majorité de 155.

D'autre part, si l'effet de la performance est inclus dans le modèle, le nombre de sièges conservateur prévu est de 157 ou 160 (selon le modèle), au-dessus du chiffre de la majorité. Le modèle de Rosenthal a pu prévoir la majorité conservatrice à

partir de son modèle de base ajusté par l'ajout de l'effet de « performance ». Cependant, certaines limites peuvent être observées :

- Le modèle n'intègre pas les stratégies de campagne (efforts, concentration des ressources...),
- Le modèle ne prend pas en considération la variation de la tendance de vote par région (le Bloc fort au Québec, libéraux forts en Ontario),
- Le modèle repose sur l'ajustement des résultats des élections précédentes avec les données de sondages, mais ne prend pas en considération les facteurs qui peuvent influencer le vote et ainsi modifier les résultats des élections. Les modèles que nous avons élaborés prennent en considération certains de ces facteurs (facteurs sociaux économiques).

1.2.2.4 Application au niveau du Québec

Le modèle 1 de notre sujet de recherche est une application de la fonction de base du modèle de Rosenthal au niveau du Québec. Nous allons appliquer les modèles 1 et 2 avec et sans l'effet de performance et nous allons les comparer avec le résultat réel.

Dans cette partie, nous avons pris connaissance des différents modèles que les chercheurs ont mis en place pour tenter de prévoir les élections présidentielles américaine ou la majorité conservatrice au niveau du Canada. Nous avons mentionné également que la conjoncture économique est le facteur le plus utilisé dans les modèles de prévision. Nous allons voir dans la section suivante les autres facteurs qui peuvent être pris en considération pour expliquer le comportement et le choix des électeurs.

1.3 Les facteurs qui influencent le vote

Selon les études menées sur le comportement des électeurs, il existe plusieurs facteurs qui influencent le vote. Bélanger et Nadeau, (2009) mettent l'accent sur deux types de facteurs : les facteurs à long terme (ou variables lourdes) et les facteurs à court terme.

Les variables lourdes ou forces à long terme sont formées par les facteurs sociaux-démographiques, qui permettent d'expliquer en partie la stabilité du comportement des électeurs. Les facteurs à court terme peuvent changer d'un scrutin à un autre. Ils regroupent les différents enjeux débattus lors de la campagne électorale ainsi que les événements précédant les élections.

1.3.1 Les facteurs sociaux démographiques

L'étude de la population québécoise a permis de distinguer des clivages importants : âge, sexe, éducation, langue, revenu, région et religion. D'après les résultats des études menées sur les élections de 2007 et 2008, Bélanger et Nadeau (2009) dressent le profil des électeurs :

« L'électeur péquiste est typiquement francophone, âgé de plus de 35 ans, moins scolarisé que l'électeur de QS ou du PVQ, a un revenu moyen, est moyennement pratiquant et habite surtout à l'extérieur des grands centres urbains (à l'exception de l'est de l'île de Montréal). »

« L'électeur Libéral est beaucoup moins francophone, un peu plus âgé, nettement plus fortuné et pratiquant, et réside surtout dans la région montréalaise. »

« L'électeur adéquiste est francophone, d'âge moyen, moins scolarisé et il habite la couronne montréalaise (Laval, Lanaudière, Laurentides et Montérégie) ou les régions du Centre-du-Québec ou de la Capitale-Nationale. L'écart entre les hommes et les femmes

n'est pas marqué dans l'électorat de l'ADQ, qui se caractérise par un degré de pratique religieuse moins élevé que chez les libéraux, mais plus intense que chez les électeurs des autres partis.»

« Les électeurs des verts sont plus anglophones, un peu plus jeunes, significativement plus scolarisés, moins pratiquants et se retrouvent davantage en milieu urbain. »

« Les électeurs de Québec solidaire sont plus concentrés à Montréal, plus scolarisés, moins fortunés, un peu plus jeunes et moins pratiquants que les partisans de l'ADQ.»

1.3.2 Les facteurs à court terme

Les enjeux débattus lors des campagnes électorales ainsi que la conjoncture économique actuelle influencent d'une manière significative le choix des électeurs.

1.3.2.1 Les enjeux

En 2008, la situation économique a constitué l'enjeu central du PLQ qui affirmait que le PLQ est le parti politique le plus compétent à gérer la crise financière (Bélanger et Nadeau).

La question nationale est l'enjeu le plus débattu, surtout lors des dernières élections. Le Parti Québécois a tenté d'obtenir les voix des partisans de la souveraineté du Québec.

La question nationale et la conjoncture économique ne constituent pas les seuls enjeux qui peuvent expliquer le choix des électeurs : en 2012, l'accessibilité au système de santé, la lutte contre la corruption et la baisse des taxes et impôts ont été les principaux enjeux (Léger Marketing, 2012).

1.3.2.2 L'image des chefs

L'image des chefs a joué un grand rôle surtout lors des dernières élections. En 2007, Mario Dumont, le chef de l'ADQ, avait obtenu une grande popularité envers les électeurs alors que les chefs des partis traditionnels ont connu une impopularité remarquable suite à l'insatisfaction envers le gouvernement Charest et le peu d'attrait du chef Péquiste André Boisclair. Ce qui explique la montée de l'ADQ lors de ces élections.

En 2008, la popularité de Jean Charest et Mario Dumont a changé. Les difficultés éprouvées par l'ADQ en tant qu'opposition ont affecté l'image du chef adéquistes. Selon les sondages, Pauline Marois avait projeté une image plus favorable que son prédécesseur André Boisclair.

1.3.2.3 La conjoncture économique

Plusieurs études affirment qu'il existe une relation entre la conjoncture économique d'un pays et le vote. Elles démontrent qu'aux États-Unis, d'une manière générale, la situation économique influence de manière significative les résultats des élections (Auberger, 2004). Cependant, cette influence est moins nette si le président sortant ne se représente pas lors des prochaines élections.

Au Canada, les études confirment qu'il existe un vote économique au niveau national : les canadiens accordent une importance significative à leur situation financière personnelle et la conjoncture économique (Godbout et Bélanger, 2002). Au Québec, Guérin et Nadeau (1998) affirment qu'il existe également un vote économique et que les électeurs sont plus sensible au problème de chômage. Par contre, Godbout et Bélanger (2002) ont affirmé que l'impact de la conjoncture économique sur le comportement des électeurs peut être affaiblie par certains facteurs tels le régionalisme, la dimension linguistique et les préférences constitutionnelles.

Dans cette partie, nous avons fait un survol des différents facteurs qui peuvent influencer le vote, en distinguant entre les forces à long terme et les variables à court terme. Ces facteurs ont permis de comprendre le comportement des électeurs et de dresser leur profil.

Dans les modèles de prévisions que nous avons établis, nous avons pris en considération les différents facteurs qui peuvent influencer le vote et intégré ces éléments dans ces modèles.

Chapitre 2 : Méthodologie de recherche

Nous allons présenter dans ce deuxième chapitre une description détaillée des données utilisées dans notre étude et la méthodologie de traitement des données. Nous allons présenter chacun des quatre modèles que nous avons élaborés afin de faire des projections des élections provinciales pour les années 2008 et 2012.

Nous allons également présenter dans un second lieu les principes de base de la prévision ainsi qu'une brève description des étapes de la régression linéaire multiple, puisque nous allons l'utiliser dans deux de nos modèles de prévision.

2.1 Les données de l'étude

2.1.1 Les données utilisées

La collecte de données s'est principalement réalisée sur internet. Les principales sources de données sont : le site du Directeur Général des Élections, Radio Canada et Leger Marketing. Nous nous sommes basé sur les modèles de prévisions étudiés dans le chapitre 1 de la revue de littérature pour collecter les données qui pourront être utilisées dans nos modèles.

Résultat des élections précédentes

Nous avons remarqué dans les différentes études menées pour construire les modèles de prévisions des élections (Section 1.2.1) que la variable la plus présente dans les modèles de prévision des élections est celle des résultats de l'élection qui précède celle que nous voulons prévoir.

Nous avons collecté pour chacune des 125 circonscriptions le nombre de votes par parti politique pour les années 2007 et 2008. Ensuite, nous les avons transformés en pourcentage de votes détenus par chaque parti politique.

Résultats de sondages préélectoraux

Les données relatives aux intentions de vote quelques jours avant les élections sont disponibles par régions administratives. Nous avons attribué le même pourcentage des intentions de vote pour les circonscriptions appartenant à la même région.

Les données sur le candidat

Nous avons appliqué la notion du président sortant de Abramowitz (1988) (section 1.2.1) et nous avons déterminé pour chacune des circonscriptions si le député actuel est candidat pour les prochaines élections dans la même circonscription et pour le même parti politique.

Les données économiques

Étant donné que la situation économique d'un pays a un impact sur les résultats des élections (Section 1.3.2), nous avons intégré la variation du PIB réel annuel ainsi que le taux de chômage une année précédant la date des élections.

Les données relatives à l'économie québécoise sont disponibles par région, nous avons appliqué le même principe que les sondages en utilisant les mêmes données pour les circonscriptions appartenant à la même région administrative.

Données sociaux démographiques

Afin d'intégrer l'impact des variables socio démographiques (variables lourdes) sur le résultat des élections, (Section 1.3.1), nous avons utilisé les données de recensement de 2006 disponible par circonscription à savoir :

- L'âge moyen dans la circonscription,
- La proportion des hommes et des femmes (sexe),
- La proportion des immigrants et des non immigrants (statut),
- Le revenu médian,

- La proportion de la population selon la connaissance du français et de l'anglais (langue),
- La proportion de la population selon le niveau de scolarité.

2.1.2 Description des variables

Le tableau en annexe 2 présente une description de toutes les variables utilisées dans les modèles de prévision ainsi que le type et les modalités de chacune.

2.1.3 Préparation des données

Nous avons structuré les variables décrites précédemment afin d'obtenir des fichiers de données pour chaque année d'élections (2007, 2008 et 2012). Les élections de 2007 vont servir comme référence pour construire les modèles prédictifs des élections de 2008, puis nous allons reprendre les données de 2008 mais cette fois-ci comme année de référence pour la prévision des élections de 2012.

Dans les modèles que nous avons étudiés, les observations sont composées des 125 circonscriptions de la province du Québec pour la prévision des élections de 2008 et 112 circonscriptions pour la prévision de celles de 2012. Nous avons éliminé 13 circonscriptions puisqu'il y avait des changements majeurs dans les délimitations en 2011.

Après avoir présenté les différentes variables et données, nous allons mettre l'accent sur les quatre modèles de prévision des élections que nous allons construire.

2.2 Description des modèles de prévision

Nous avons construit quatre modèles de prévision pour les élections québécoises. Les modèles sont indépendants et n'utilisent pas forcément les mêmes variables. L'objectif reste le même : estimer le nombre de sièges pour chacun des principaux partis politiques. Nous allons tenter dans un premier temps d'appliquer le modèle

de base de Rosenthal au niveau provincial afin de prévoir le nombre de sièges que pourra détenir les principaux partis politiques lors des élections de 2008 et 2012. Le deuxième modèle a pour objectif de savoir si les sondages préélectorales peuvent prévoir seuls les résultats des élections. Ce modèle basé sur les sondages transforme l'intention de vote en pourcentage de vote. Le troisième modèle a pour objectif de prévoir le pourcentage de vote de chaque parti politique dans chacune des circonscriptions en intégrant plusieurs variables qui peuvent influencer le vote (résultat des élections précédentes, les résultats de sondages, les facteurs à long terme et les variables économiques). Le quatrième modèle a pour objectif d'établir des prévisions sur l'erreur de sondage. Au lieu de prédire le pourcentage de vote, nous allons calculer l'erreur des sondages préélectorales. Ce modèle permet d'ajuster les sondages réalisés avant les élections et obtenir les résultats du nombre de sièges par parti politique.

Les quatre modèles que nous allons étudier dans cette partie sont donc les suivants :

Modèle 1 : Modèle de base de Rosenthal

Modèle 2: Modèle basé uniquement sur les sondages

Modèle 3 : Modèle de prévision du pourcentage de vote

Modèle 4 : Modèle de prévision de l'erreur de sondage

2.2.1 Le modèle de base de Rosenthal

Nous avons utilisé la fonction de base de Rosenthal pour prédire le nombre de sièges que va détenir chacun des partis politiques du Québec lors des élections de 2008 et 2012. Pour chacune des élections, nous avons appliqué la fonction de vote du modèle de Rosenthal sur chacune des circonscriptions du Québec. Nous avons utilisé les deux modèles proposés par Rosenthal sans et avec effet de performance (section 1.2.2). Les résultats sont présentés dans le Chapitre 3.

2.2.2 Le modèle basé uniquement sur les sondages :

L'objectif de ce deuxième modèle est de savoir si les sondages peuvent prévoir non seulement les intentions de vote, mais également le nombre de sièges que peut détenir chaque parti politique lors des élections.

Pour cela, nous avons transformé les intentions de vote résultant des derniers sondages en pourcentage de vote estimé pour chaque parti politique par circonscription. Nous avons donc multiplié le pourcentage des intentions de vote de chaque parti politique par le nombre de votes par circonscription lors des dernières élections. La fonction est la suivante :

$$V = P * T.$$

- **Variable à expliquer :**

V= Nombre de votes du parti politique dans une circonscription

- **Variables explicatives :**

P = pourcentage des intentions de vote pour le parti politique (résultats des derniers sondages pré-électoraux)

T = nombre total de votes d'une circonscription lors de la précédente élection.

Nous avons appliqué ce modèle pour les élections de 2008 et 2012. Par la suite nous avons comparé les résultats de chaque circonscription avec le résultat réel. Les résultats seront présentés dans le Chapitre 3.

2.2.3 Le modèle de prévision du pourcentage de vote

Le but de ce troisième modèle est de prévoir le pourcentage des trois principaux partis politiques pour chacune des circonscriptions. Nous avons utilisé la régression linéaire multiple puisque nous avons une variable à expliquer (pourcentage de vote) et plusieurs variables explicatives,.

La fonction de vote est donc la suivante :

$$\begin{aligned}
 Y_{\text{parti}} = & \beta_0 + (\beta_1 * R_{\text{parti}}) + (\beta_2 * P_{\text{parti}}) + (\beta_3 * \text{candidat}) + (\beta_4 * \text{masculin}) \\
 & + (\beta_5 * \text{langue2}) + (\beta_6 * \text{langue3}) + (\beta_7 * \text{langue4}) + (\beta_8 * D_1) \\
 & + (\beta_9 * D_2) + (\beta_{10} * D_3) + (\beta_{11} * D_4) + (\beta_{12} * D_5) \\
 & + (\beta_{13} * \text{age_moyen}) + (\beta_{14} * \text{immigrant}) + (\beta_{15} * \text{taux}_{\text{chomage}}) \\
 & + (\beta_{16} * \text{revenu_median}) + (\beta_{17} * \text{PIB}).
 \end{aligned}$$

Pour construire le modèle, nous avons utilisé le logiciel SAS pour réaliser la régression linéaire et intégrer toutes les variables qui peuvent influencer le vote. Nous avons intégré dans notre modèle toutes les variables susceptibles d'influencer le vote. Mais puisque nous disposons de plusieurs variables, nous avons sélectionné pour chaque parti politique les variables significatives. La démarche statistique est présentée dans la section 2.4.

Nous avons par la suite tenté d'améliorer notre modèle en intégrant les interactions entre les variables (scénario 2) :

- Interactions entre les résultats des élections et les différents facteurs sociaux-économiques,
- interactions entre les résultats des élections et le candidat présenté dans la circonscription,
- interactions entre les résultats des élections et les résultats de sondages,
- Interactions entre les résultats des sondages et les différents facteurs sociaux-économiques,

- interactions entre les résultats des sondages et le candidat présenté dans la circonscription,

La liste des interactions est présentée dans l'annexe 3. Les résultats des différents modèles de prévisions des élections de 2008 et 2012 sont comparés avec les résultats réels des élections afin de déduire la pertinence de notre modèle. Les résultats des 2 scénarios (avec sélection de variables et avec interactions) sont présentés dans le Chapitre 3.

2.2.4 Modèle de prévision de l'erreur de sondage :

Le quatrième modèle a pour objectif d'estimer la marge d'erreur des sondages pré-électoraux. Avec cette estimation, nous avons ajusté les sondages pré-électoraux et nous avons établi une prévision des résultats des élections. L'erreur de sondage est calculée de la manière suivante :

Erreur de sondage

$$= \text{résultat du sondage des élections par circonscription} \\ - \text{résultat réel des élections.}$$

Dans notre modèle de régression, nous avons utilisé les mêmes variables explicatives que le modèle 3. Nous avons cependant éliminé les variables de sondage puisque c'est la variable que nous allons ajuster. La fonction de régression de chaque parti politique est la suivante :

$$Y(\text{Erreur } n) = \beta_0 + (\beta_1 * \text{Erreur } n - 1) + (\beta_2 * \text{candidat}) + (\beta_3 * \text{masculin}) \\ + (\beta_4 * \text{langue2}) + (\beta_5 * \text{langue3}) + (\beta_6 * \text{langue4}) + (\beta_7 * D_1) \\ + (\beta_8 * D_2) + (\beta_9 * D_3) + (\beta_{10} * D_4) + (\beta_{11} * D_5) + (\beta_{12} \\ * \text{age_moyen}) + (\beta_{13} * \text{immigrant}) + (\beta_{14} * \text{taux_chomage}) + (\beta_{15} \\ * \text{revenu_median}) + (\beta_{16} * \text{PIB})$$

Erreur_n : Erreur de sondage à prévoir (Variable à expliquer)

Erreur_{n-1} : Erreur de sondage calculé à partir des élections précédentes (variable explicative)

À partir de ce modèle, nous pouvons corriger les sondages des élections afin d'obtenir les prévisions finales.

Nous avons élaboré également deux scénarios comme dans le modèle précédent.

Nous avons gardé les mêmes interactions que dans le modèle 3 mais en remplaçant celles qui contiennent les résultats de sondages et des dernières élections par l'erreur de sondages.

Avant de présenter les résultats des quatre modèles de prévision, nous allons présenter les principes de base de la prévision des élections et nous allons décrire la démarche statistique adoptée pour les modèles 3 et 4.

2.3 Principes de base de la prévision

Avant de bâtir des modèles de prévision pour des élections, un chercheur doit prendre en considération les éléments suivants :

a) La neutralité (non appartenance à un parti politique)

En tant que scientifique, il est important d'avoir une image non partisane. Le fait d'appartenir à un parti politique peut affecter la crédibilité du chercheur.

b) L'incertitude

Les prévisions statistiques pour les futures élections sont toujours incertaines, les modèles sont plutôt stochastiques que déterministes. Même si le modèle est correctement défini, il y aura de la variabilité dans les estimations des paramètres. Une autre source d'erreur provient de la mesure des variables, en particulier les données de sondages, avec l'erreur d'échantillonnage intégrée.

L'erreur est attendue dans ce genre de recherche. L'objectif du chercheur est de la réduire au minimum (Lewis-Beck, 2005).

c) L'évaluation des prévisions

Selon Lewis-Beck (2005), un instrument de prévision doit être évalué selon quatre critères : la précision, l'avance (lead), la parcimonie et la reproductibilité.

Précision : en prévision, la norme d'évaluation est l'exactitude. Si un modèle produit toujours des prévisions inexactes, il est probablement erroné.

Lead : le modèle de prévision est pertinent s'il est établi avant l'évènement. Plus le modèle peut être appliqué longtemps avant les élections, mieux c'est.

Parcimonie : toutes choses égales par ailleurs, l'idéal est de retenir le modèle le plus simple qui donne un ajustement satisfaisant. Avec trop de paramètres, le modèle manque de robustesse.

2.4 Démarches statistiques

Dans les modèles 3 et 4, nous avons opté pour la régression linéaire multiple pour la prévision des élections. Dans cette partie, nous allons présenter les objectifs de la régression linéaire, la fonction de régression, puis nous allons expliquer la méthode de sélection des variables que nous avons choisie ainsi que les critères de sélection. Nous allons par la suite expliquer comment nous avons procédé à la comparaison des résultats de prévision avec le résultat réel des élections.

2.4.1 La régression linéaire multiple

La régression linéaire est utilisée afin d'expliquer le comportement d'une variable quantitative Y (variable dépendante) en utilisant un modèle basé sur P variables explicatives X (variables indépendantes). La régression linéaire permet également de modéliser et faire des projections (Larocque, 2010).

Les buts d'une analyse de régression sont :

- a- Comprendre comment et dans quelles mesures les variables indépendantes X influencent la variable dépendante Y . On pourra quantifier l'influence de ces variables et tester des hypothèses quant à ces effets.
- b- Prévision : développer un modèle pour prévoir des valeurs de Y futures à partir des variables X .

Dans notre étude, nous avons établi une prévision de la valeur de Y à partir de plusieurs variables indépendantes. Nous avons utilisé la régression linéaire multiple. Le modèle de régression linéaire multiple est formulé comme suit :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i \quad i = 1, \dots, n \quad .$$

Afin de tester la validité de nos modèles de régression, nous avons calculé le coefficient de détermination ajusté (R^2_{adj}) et l'erreur quadratique (erreur de prévision).

Ces deux mesures nous ont permis de comparer les différents modèles et choisir le modèle le plus pertinent, qui se rapproche le plus des résultats réels des élections.

2.4.2 La sélection de variables

La recherche des facteurs explicatifs dans nos modèles de régression a été effectuée avec la procédure GLMSELECT du logiciel SAS (version 9.2), au moyen d'une procédure itérative de type « stepwise », et du critère de sélection AIC.

a. La procédure GLMSELECT :

C'est une méthode expérimentale introduite par SAS dans la version 9.2 (SAS Institute, 2008). La procédure de GLMSELECT effectue une sélection de variables dans le cadre de modèles de régression linéaire. La procédure offre des capacités étendues de personnalisation de la sélection avec une grande variété de critères de sélection et d'arrêt. La procédure GLMSELECT permet de faire une recherche de

type « stepwise » en prenant des critères autres que des p-values pour faire entrer, sortir et choisir le modèle final.

La procédure GLMSELECT est un croisement entre deux procédures SAS pour les modèles linéaires: REG et GLM. La procédure GLMSELECT vient corriger certaines limites de ces deux procédures³.

b. Méthodes de sélection de variables

Dans notre étude, on dispose de plusieurs variables explicatives; nous avons tenté d'introduire toutes les variables susceptibles d'avoir un effet significatif sur nos prévisions des élections.

Par conséquent, nous avons affiné nos modèles de régression multiple de départ (incluant toutes les variables) en effectuant une sélection de variables. La sélection de variables s'avère nécessaire également parce que le nombre d'observations est faible par rapport au nombre de variables explicatives.

Il existe plusieurs techniques de sélection de variables qui permettent de sélectionner un nombre réduit de variables qui expliquent pourtant une quantité raisonnable de variation .

La méthode Backward : Elle consiste à intégrer toutes les variables explicatives, et à retirer une par une les variables qui sont les moins significatives. Dans cette méthode, une fois qu'une variable est sortie du modèle, elle ne peut plus y entrer.

La méthode Forward : C'est l'inverse de la méthode Backward; elle sélectionne en premier lieu la variable explicative la plus corrélée puis elle sélectionne, parmi celles qui restent, la variable explicative dont la corrélation partielle est la plus élevée. Et ainsi de suite tant qu'il reste des variables dont le coefficient de corrélation partiel est significatif. Une fois qu'une variable est dans le modèle, elle y reste.

La méthode Stepwise : C'est une combinaison des deux méthodes, elle consiste à faire entrer les variables l'une après l'autre dans le modèle (selon leur corrélation partielle) par sélection progressive et, à chaque étape, à vérifier si les corrélations

³ <http://www.association-assq.qc.ca/2012/05/24/la-chronique-sas-%C2%ABdiviser-pour-regner-avec-proc-glmselect%C2%BB/>

partielles de l'ensemble des variables déjà introduites sont encore significatives (une variable qui ne le serait plus serait rejetée). Avec la méthode «stepwise», une variable peut entrer dans le modèle et sortir plus tard dans le processus.

Par conséquent, parmi les trois méthodes, la méthode « Stepwise » est généralement préférable aux méthodes « Forward » et « Backward », car elle inspecte potentiellement un plus grand nombre de modèles. Dans notre étude, nous avons utilisé la méthode « Stepwise » pour la sélection des variables dans les modèles de régression.

c. Les critères de sélection des variables

Les critères AIC⁴ et BIC⁵ peuvent être utilisés comme outils de sélection de variables en régression linéaire.

Le critère AIC permet de pénaliser les modèles en fonction du nombre de paramètres afin de satisfaire le critère de parcimonie. Pour le critère BIC, à la différence du critère AIC, la pénalité dépend de la taille de l'échantillon et pas seulement du nombre de paramètres. En fait, ces deux critères découlent de la méthode d'estimation basée sur le maximum de vraisemblance. Pour chacun de ces critères, plus la valeur est petite, mieux c'est. On peut utiliser les critères AIC ou BIC pour chercher le « meilleur sous ensemble » de variables. Avec la procédure GLMSELECT, il est possible de choisir la variable à entrer ou à retirer, à chaque étape d'une procédure stepwise, avec ces critères au lieu d'utiliser les p-values des corrélations partielles.

L'AIC est un excellent critère pour comparer différents modèles de prévisions, puisque dans notre cas, notre objectif n'est pas déterminer si les variables sont significatives.

⁴ « Akaike Information Criterion »

⁵ « Bayesian Information Criterion »

2.4.3 Comparaison des prévisions avec le résultat réel

Nous avons utilisé le tableau de contingence afin de compter le nombre de sièges correctement prédit par nos modèles de prévision et obtenir le taux de bonne classification pour chacun des modèles. Une matrice de confusion ou tableau de contingence sert à évaluer la qualité d'une classification. Elle est obtenue en comparant les données classées avec des données de référence (résultat réel des élections).

2.4.4 Les indicateurs de performance

Trois indicateurs vont mesurer la performance de nos modèles :

Le taux de bonne classification ou taux de bonne prédiction : calculé à partir du tableau de contingence.

L'erreur quadratique ou **erreur de prévision** qui mesure la moyenne de l'erreur quand on utilise notre modèle pour l'ensemble de la population. Plus l'erreur est petite, meilleur est l'ajustement du modèle. L'erreur de prévision est estimée à travers la fonction suivante :

$$TMSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{f}(X_{1i}, \dots, X_{pi}))^2 .$$

Le R^2_{adj} ou coefficient de détermination ajusté qui permet de mesurer la force de la relation linéaire entre la variable à expliquer (Y) et les variables explicatives (X). Plus le R^2 ajusté est grand, plus il existe une relation linéaire forte entre X et Y (Larocque, 2006) Le R^2 ajusté peut être utilisé comme outil de sélection de modèle, plus la valeur de R^2_{adj} est élevée, meilleur est le modèle.

Comme nous l'avons expliqué dans la section 2.1.3, les modèles sont estimés avec les données de l'élection précédente (apprentissage) à celle que l'on veut prédire.

Ainsi, le taux de bonne classification et l'erreur de prévision sont calculés en utilisant l'échantillon que l'on veut prédire (test).

Après avoir décrit les quatre modèles statistiques que nous avons utilisés, nous allons présenter dans le troisième et dernier chapitre les résultats obtenus. Nous allons par la suite comparer les résultats et choisir le modèle qui se rapproche le plus des résultats réels des élections. Nous allons pouvoir également déterminer si nos modèles de prévision des pourcentages de votes et la prévision de l'erreur de sondage performant mieux que le modèle de Rosenthal.

Chapitre 3 : Analyses et résultats

Dans la première section de ce présent chapitre, nous allons présenter les résultats de chaque modèle étudié, puis par la suite nous allons présenter un tableau récapitulatif des résultats afin de pouvoir comparer ces modèles et déduire le meilleur modèle de prévision des élections.

3.1 Résultat du modèle de base de Rosenthal :

Nous avons appliqué le modèle de base de Rosenthal au niveau provincial (les deux fonctions). Nous avons inclus également l'effet de performance des partis politiques. Les tableaux suivants présentent le nombre de sièges bien prédits pour chaque modèle ainsi que les résultats réels des élections :

2008	modèle 1		modèle2		résultat des élections
	sans performance	avec performance	sans performance	avec performance	
PLQ	52	47	52	47	66
PQ	28	27	29	30	51
ADQ	0	3	0	5	7
autre	0	0	0	0	1
Total	80	77	81	82	125

Tableau 3:Résultat du modèle de base de Rosenthal pour la prévision des élections provinciales de 2008.

2012	modèle 1		Modèle 2		résultat des élections
	sans performance	avec performance	sans performance	avec performance	
PLQ	20	20	19	22	49
PQ	38	35	37	37	47
CAQ	8	9	8	9	14
autre	0	0	0	0	2
total	66	64	64	68	112

Tableau 4: Résultat du modèle de base de Rosenthal pour la prévision des élections provinciales de 2012.

On remarque que dans les deux élections, les résultats du modèle 2 avec l'effet de performance se rapprochent le plus du résultat réel des élections.

Nous avons par la suite totalisé le nombre de sièges que nous avons correctement prédit avec le nombre de sièges total. Le taux de bonne classification est le suivant :

	modèle 1		modèle2	
	sans performance	avec performance	sans performance	avec performance
2008	64%	62%	64%	65%
2012	59%	57%	57%	61%

Tableau 5: Taux de bonne classification du modèle de base de Rosenthal

On remarque que dans le modèle 2, les résultats avec l'effet de performance sont légèrement meilleurs que les autres modèles. Pour les deux élections, le modèle 2 a pu prévoir le parti politique qui forme le gouvernement.

Afin de comparer le modèle de Rosenthal avec les autres modèles de prévision, nous avons pris en considération le modèle 2 avec effet de performance qui semble le plus performant par rapport aux autres modèles.

3.2 Résultat du modèle basé uniquement sur les sondages :

Nous avons transformé les intentions de vote en prévision du nombre de sièges détenus par chacun des principaux partis politiques à partir de la fonction décrite dans le chapitre 2 point 2.2.2. Les résultats sont les suivants :

	2008		2012	
	Résultat de la prévision	résultat des élections	Résultat de la prévision	résultat des élections
PLQ	54	66	19	49
PQ	25	51	38	47
ADQ/CAQ	0	7	8	14
autre	0	1	0	2
Total	79	125	65	112

Tableau 6: Nombre de sièges bien prédits pour le modèle basé uniquement sur les sondages

On constate à partir des résultats que le modèle parvient à prévoir correctement 63% des sièges pour les élections de 2008 et 58% pour celles de 2012.

3.3 Résultat du modèle de prévision du pourcentage de vote

L'application de la régression linéaire (code SAS en annexe 5) pour la prévision du pourcentage de vote pour les élections de 2008 et 2012 a fait ressortir les résultats suivants.

3.3.1 Les variables significatives

Les résultats des élections précédentes ainsi que les résultats de sondages sont présents dans la majorité des cas. On peut les considérer comme les variables explicatives les plus importantes. Par contre, les variables qui peuvent influencer le vote et donc le résultat des élections diffèrent selon le parti politique et selon les élections. En général, les facteurs les plus récurrents sont : la situation économique de la province, le niveau scolaire et le sexe des électeurs. L'Annexe 4 présente le détail de la sélection de variables.

3.3.2 Prévision du nombre de sièges :

Nous avons comparé le nombre de sièges « bien prédits » avec les résultats réels des élections. Les résultats se présentent comme suit :

2008	Scénario 1	Scénario 2	Résultat réel
PLQ	40	38	66
PQ	25	25	51
ADQ	6	7	7
Autre	0	0	1
Total	71	70	125

Tableau 7: Modèle de prévision du pourcentage de vote: Prévision du nombre de sièges par parti politique pour les élections de 2008

2012	Scénario 1	Scénario 2	Résultat réel
PLQ	45	43	49
PQ	31	35	47
CAQ	0	0	14
Autre	0	0	2
Total	76	77	112

Tableau 8: Modèle de prévision du pourcentage de vote: Prévision du nombre de sièges par parti politique pour les élections de 2012

Le modèle prédit correctement 56% du nombre de sièges en 2008 et 68% en 2012. L'ajout des interactions n'a pas une incidence significative sur les résultats du modèle. Malgré que certaines interactions semblent significatives dans le modèle de régression (annexe 4), le nombre de sièges bien prédits reste quasiment le même que dans le premier scénario. On remarque également que le scénario 2 n'intègre pas les variables explicatives principales (résultats des élections précédentes et résultats des sondages). Le modèle n'a pas pu prévoir un gouvernement Péquiste en 2012 contrairement aux deux modèles précédents.

3.4 Résultat du modèle de prévision de l'erreur de sondage

L'application de la régression linéaire (code SAS en annexe 5) pour la prévision de l'erreur de sondages pour les élections de 2008 et 2012 a fait ressortir les résultats suivants.

3.4.1 Variables significatives

Comme pour le troisième modèle, les variables significatives changent d'un parti à un autre et d'une élection à une autre. En général, les variables les plus présentes dans ce modèle sont : l'erreur de sondages des élections précédentes, le niveau d'études et l'âge moyen. Les tableaux en annexe 4 présentent en détail les variables significatives pour chaque scénario.

3.4.2 Prévision du nombre de sièges

Les résultats du nombre de sièges « Bien prédits » sont les suivants :

2008	Scénario 1	Scénario 2	Résultat réel
PLQ	46	42	66
PQ	44	47	51
ADQ	3	2	7
Autre	0	0	1
Total	93	91	125

Tableau 9:Modèle de prévision de l'erreur de sondages: Prévision du nombre de sièges par parti (2008)

2012	Scénario 1	Scénario 2	Résultat réel
PLQ	24	23	49
PQ	40	41	47
ADQ	9	9	14
Autre	0	0	2
Total	73	73	112

Tableau 10: Modèle de prévision de l'erreur de sondages: Prévision du nombre de sièges par parti (2012)

Le 4^{ème} modèle prédit correctement 74% du nombre de sièges en 2008 et 65% en 2012. Comme le modèle précédent, les interactions n'ont pas d'effet significatif sur le résultat global et n'intègre pas la variable explicative principale (erreur de sondages des élections précédentes). Par contre, le modèle a pu prévoir un gouvernement péquiste pour les élections de 2012.

3.5 Tableau récapitulatif des résultats des 4 modèles

Nous allons présenter dans ce chapitre la synthèse des résultats des modèles de prévisions pour les élections de 2008 et 2012. Les annexes 6a et 6b présentent les résultats détaillés pour les modèles 3 et 4 : l'estimation des paramètres, les variables significatives ainsi que le tableau de contingence.

1. Prévision des élections de 2008 :

		TBC	R2 ajusté*	AIC	BIC	TMSE
modèle 1	Rosenthal	65%	-	-	-	0.0285
modèle 2	Sondage	63%	-	-	-	0.025
modèle 3	Pourcentage vote					
Scénario 2	Avec sélection	57%	73%	-545.53	-653.67	0.0195
Scénario 3	Avec interaction	56%	74%	-551.81	-661.84	0.0219
modèle4	Erreur sondage					
Scénario 2	Avec sélection	74%	51%	-531.9	-647.59	0.0097
Scénario 3	Avec interaction	73%	57%	-542.56	-647.87	0.0114

* moyenne des 3 partis politiques

Tableau 11: Tableau récapitulatif des résultats des 4 modèles pour les élections de 2008

2. Pr vision des  lections de 2012

		TBC	R2 ajust�*	AIC	BIC	TMSE
mod�le 1	Rosenthal	61%	-	-	-	0.01833
mod�le 2	Sondage	58%	-	-	-	0.0149
mod�le 3	Pourcentage vote					
Sc�nario 2	Avec s�lection	68%	70%	-486.79	-589.92	0.0287
Sc�nario 3	Avec interaction	70%	75%	-503.45	-595.71	0.0368
mod�le4	Erreur sondage					
Sc�nario 2	Avec s�lection	65%	64%	-509.93	-613.97	0.0167
Sc�nario 3	Avec interaction	65%	70%	-529.48	-617.95	0.0169

* moyenne des 3 partis politiques

Tableau 12: Tableau r capitulatif des r sultats des 4 mod les pour les  lections de 2012

3.6 Analyse des r sultats des mod les de pr vision

Pour les  lections de 2008

Lorsque nous comparons le taux de bonne classification des quatre mod les, nous concluons que le mod le de pr vision de l'erreur de sondages a une meilleure performance que les autres mod les malgr  que les variables expliquent seulement 51% de la variabilit  du pourcentage de votes. Le mod le a pu correctement pr dire jusqu'  74% du nombre de si ges d tenus par chaque parti politique et aussi le parti politique qui forme le gouvernement. Comme nous l'avons expliqu  auparavant, les interactions entre les variables n'ont pas un impact significatif sur le r sultat du mod le, nous n'allons pas prendre en consid ration le sc nario 2.

Pour les  lections de 2012

En utilisant le m me crit re de performance (taux de bonne classification), nous avons constat  que le mod le de pr vision du pourcentage de votes est le mod le le plus performant. Ce mod le a permis de pr dire correctement jusqu'  70% des si ges. Les variables utilis es expliquent jusqu'  75% de la variabilit  du pourcentage de votes. Le sc nario 2 qui int gre les interactions entre les variables est  galement

mis à l'écart. Par contre, ce modèle n'a pas pu prévoir le parti qui forme le gouvernement en 2012.

En se basant sur le taux de bonne classification, le modèle de prévision de l'erreur de sondages est plus performant que les autres modèles pour prévoir les résultats des élections de 2008 alors que pour celles de 2012, c'est le modèle de prévision du pourcentage de votes qui performe le mieux.

Or, une des règles essentielles de la prévision est d'avoir un modèle unique qui va permettre de prévoir des élections futures. Afin de faire le choix entre ces deux modèles, nous avons comparé les valeurs de l'erreur de prévision pour les deux élections et nous avons constaté que l'erreur est moins élevée pour le modèle 4.

Nous pouvons donc déduire que le modèle de prévision de l'erreur de sondages est le plus robuste et donc le plus susceptible à prévoir les résultats des futures élections. Ainsi, à l'approche des élections, le modèle va ajuster les sondages et va permettre d'estimer le nombre de sièges détenus par chaque parti politique et prédire le parti qui va former le prochain gouvernement.

Conclusion

La prévision des élections dépend de plusieurs facteurs. Notre but était d'utiliser certains de ces facteurs afin de construire un modèle prédictif qui estime le nombre de sièges que va détenir chacun des principaux partis politiques lors des prochaines élections au Québec. Suite aux différentes recherches réalisées dans ce domaine, nous avons essayé d'intégrer les variables susceptibles d'avoir un impact sur les résultats des élections.

L'application du modèle de base de Rosenthal était le point de départ de notre recherche. Puis nous avons construit d'autres modèles dans l'optique d'obtenir de meilleurs résultats. Le modèle de prévision de l'erreur de sondages s'est avéré le plus performant par rapport aux autres modèles étudiés; les résultats étant les plus proches du résultat réel des élections. Ce modèle a été construit à partir d'une régression linéaire multiple. Il intègre les variables socio-économiques, l'erreur de sondages et les résultats des élections précédentes. Il faut souligner que dans ce modèle, toutes les variables citées sont incluses; les plus significatives vont être sélectionnées par le modèle de régression.

Le modèle de prévision de l'erreur de sondages a pu prévoir jusqu'à 74% du nombre de sièges. Ce taux pourrait être amélioré si on disposait des données par circonscription et d'informations supplémentaires sur les facteurs à court terme qui influencent le comportement des électeurs. En effet, plusieurs contraintes ont fait obstacle :

- a) Les sondages préélectorales sont réalisés par une entreprise privée, nous avons seulement utilisés les données agrégées publiées par la firme. La présentation des informations est différente d'une élection à une autre et les résultats sont présentés par régions administratives et non pas par circonscription. Ce qui nous a amené à ajuster les plus possible ces informations afin de les intégrer dans nos modèles.

- b) Les données collectées des rapports de sondages permettent d'avoir une idée sur les intentions de vote des citoyens. Cependant, les données concernant les facteurs à court terme (image des chefs, les enjeux politiques, etc) ne figurent pas en détail dans ces publications.
- c) Les données de sondages et du PIB ne sont pas disponibles par circonscription, nous avons été contraints d'utiliser les données régionales. Or, celles-ci ne permettent pas d'avoir une projection précise pour une circonscription donnée. Deux circonscriptions d'une même région peuvent être très divergentes. Par exemple, la région de Montréal est constituée par 28 circonscriptions. Les circonscriptions Westmount, Mont-Royal et Outremont sont en faveur des libéraux alors que les circonscriptions de Hochelaga-Maisonneuve, Gouin et Pointe-aux-Trembles sont en faveur des péquistes. Cette divergence des intentions de votes dans les circonscriptions de la même région n'est pas prise en considération dans les sondages ainsi que les données relatives au PIB.
- d) Le redécoupage électoral de 2011 a été une contrainte pour bâtir un modèle de prévision des élections de 2012. La délimitation a été changée pour plusieurs circonscriptions. Des circonscriptions ont été nouvellement créés et d'autres supprimées. Nous avons eu l'obligation de supprimer 13 circonscriptions et ignorer les changements de délimitations apportés à certaines circonscriptions.

Le modèle de prévision de l'erreur de sondages peut être appliqué pour les prochaines élections. Le modèle va ajuster les résultats des sondages et va estimer le nombre de sièges pour chaque parti politique et pourra donc prédire qui formera le prochain gouvernement. Aussi, le modèle va intégrer les données de recensement de 2011. Nous disposerons donc des données socioéconomiques plus récentes.

Pour continuer cette recherche, il est possible de travailler avec d'autres modèles statistiques ou encore intégrer d'autres facteurs qui peuvent influencer les résultats des élections, en cherchant le plus possible les données par circonscription.

Références bibliographiques :

- Abramowitz, Alan I. (1988). An Improved Model for Predicting Presidential Election Outcomes. *PS: Political Science and Politics*, 21(4), 843-847. doi: 10.2307/420023
- Abramowitz, Alan I. (2004). When Good Forecasts Go Bad: The Time-for-Change Model and the 2004 Presidential Election. *PS: Political Science and Politics*, 37(4), 745-746. doi: 10.2307/4488899
- Alesina, Alberto, Londregan, John, & Rosenthal, Howard. (1996). The 1992, 1994 and 1996 Elections: A Comment and a Forecast. *Public Choice*, 88(1/2), 115-125. doi: 10.2307/30027256
- Auberger, Antoine. (2004). Les fonctions de vote : un survol de la littérature*. *L'Actualité Économique*, 80(1), 95-107.
- Bélanger, Éric, & Nadeau, Richard. (2009). *Le comportement électoral des Québécois Montréal : Presses de l'Université de Montréal, 2009.*
- Brody, Richard, & Sigelman, Lee. (1983). Presidential Popularity and Presidential Elections: An Update and Extension. *The Public Opinion Quarterly*, 47(3), 325-328. doi: 10.2307/2749005
- Canada, Parlement. (1994). *Les sondages d'opinion au Canada* La Direction de la Recherche Parlementaire du Canada Retrieved from <http://www.parl.gc.ca/Content/LOP/researchpublications/bp371-f.pdf>.
- Fair, Ray C. (1978). The Effect of Economic Events on Votes for President. *The Review of Economics and Statistics*, 60(2), 159-173. doi: 10.2307/1924969
- Godbout, Jean-François, & Bélanger, Éric. (2002). La dimension régionale du vote économique canadien aux élections fédérales de 1988 à 2000. *Canadian Journal of Political Science / Revue canadienne de science politique*, 35(3), 567-588. doi: 10.2307/3233115
- Guérin, Daniel, & Nadeau, Richard. (1998). Clivage linguistique et vote économique au Canada. *Canadian Journal of Political Science / Revue canadienne de science politique*, 31(3), 557-572. doi: 10.2307/3233044
- Hibbs, Douglas A., Jr. (2000). Bread and Peace Voting in U.S. Presidential Elections. *Public Choice*, 104(1-2), 149-180.
- Institute, SAS. (2008). *SAS/STAT 9.2 User's Guide The GLMSELECT Procedure* (pp. 2682).
- Kramer, Gerald H. (1971). Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964. *The American Political Science Review*, 65(1), 131-143. doi: 10.2307/1955049
- Larocque, Denis. (2006). Analyse multidimensionnelle appliquée. *Recueil 6-602-07 A07*, 2006, 408 p.
- Larocque, Denis. (2010). *Techniques quantitatives en marketing Recueil 3-670-07 A10*
- LegerMarketing. (2012). *Élections Québec 2012 Sondage électoral: Leger Marketing.*
- Lewis-Beck, Michael S. (2005). Election Forecasting: Principles and Practice. *British Journal of Politics & International Relations*, 7(2), 145-164. doi: 10.1111/j.1467-856X.2005.00178.x
- Lewis-Beck, Michael S., & Rice, Tom W. (1984). Forecasting Presidential Elections: A Comparison of Naive Models. *Political Behavior*, 6(1), 9-21. doi: 10.2307/586044
- Martin, Patrick. (1987). « Do Opinion Polls Influence Voters? » *Politics : Canada* 303.
- Martin, Pierre. (1993). "Ces sondages ventriloques" *Le Monde diplomatique*, 49.
- Rosenthal, Jeffrey S. (2011). Was the conservative majority predictable? *The Canadian Journal of Statistics*, 39(4), 721.

Annexes

Annexe 1a: liste des circonscriptions du Québec (2001)

Abitibi-Témiscamingue	Lanaudière/Laurentides	Montreal-est	Saguenay-Lac-Saint-Jean / Côte-Nord
Abitibi-Est	Argenteuil	Anjou	Papineau
Rouyn-Noranda Témiscamingue	Bertrand	Bourassa-Sauvé	Pontiac
Abitibi-Ouest	Rousseau	Jeanne-Mance—Viger	La Pinière
Bas-Saint-Laurent/Gaspésie-Îles-de-la-Madeleine	Berthier	Laurier-Dorion	Dubuc
Bonaventure	Blainville	Saint-François	Chicoutimi
Kamouraska-Témiscouata	Deux-Montagnes	Viau	Duplessis
Matane	Groulx	Laporte	Jonquière
Gaspé	Joliette	Bourget	René-Lévesque
Îles-de-la-Madeleine	La Peltrie	Gouin	Roberval
Matapédia	L'Assomption	Hochelaga-Maisonneuve	Ungava
Rimouski	Masson	Pointe-aux-Trembles	Montréal
Rivière-du-Loup	Mirabel	Rosemont	Saint-Maurice
Capitale-Nationale	Prévost	Mercier	Brome-Missisquoi
Jean-Talon	Terrebonne	Montreal-ouest	Châteauguay
Louis-Hébert	Mauricie/Centre-du-Québec	Acadie	LaFontaine
Charlevoix	Laviolette	D'Arcy-McGee	Soulanges
Taschereau	Arthabaska	Jacques-Cartier	Vaudreuil
Labelle	Maskinongé	Marguerite-Bourgeoys	Saint-Laurent
Charlesbourg	Trois-Rivières	Marquette	Lac-Saint-Jean
Jean-Lesage	Champlain	Mont-Royal	Beauharnois
Montmorency	Drummond	Nelligan	Borduas
Portneuf	Nicolet-Yamaska	Notre-Dame-de-Grâce	Marie-Victorin
Vanier	Estrie	Outremont	Richelieu
Chauveau	Mégantic-Compton	Robert-Baldwin	Taillon
Chaudière-Appalaches	Orford	Verdun	Vachon
Frontenac	Richmond	Westmount—Saint-Louis	Verchères
Beauce-Sud	Sherbrooke	Crémazie	Huntingdon
Bellechasse	Saint-Henri—Sainte-Anne	Sainte-Marie—Saint-Jacques	La Prairie
Lévis	Johnson	Outaouais	Saint-Jean
Montmagny-L'Islet	Laval	Saint-Hyacinthe	Chambly
Beauce-Nord	Chomedey	Chapleau	Iberville
Chutes-de-la-Chaudière	Fabre	Gatineau	Marguerite-D'Youville
Lotbinière	Laval-des-Rapides	Hull	Shefford
	Mille-Îles		
	Vimont		

Annexe 1b: Redécoupage de la carte électorale québécoise (2011)

Une nouvelle carte électorale est adoptée le 19 octobre 2011. Le nombre de circonscriptions reste stable à 125.

- La circonscription d'Anjou est renommée Anjou–Louis-Riel.
- Charlevoix devient Charlevoix–Côte-de-Beaupré par ajout de la partie ouest de Montmorency.
- Drummond disparaît, divisée entre la nouvelle Drummond—Bois-Francs et Johnson.
- Frontenac et une partie de Lotbinière sont regroupés pour former Lotbinière-Frontenac.
- Marguerite-D'Youville disparaît et Montarville apparaît.
- Matane disparaît, divisée entre la nouvelle circonscription de Matane-Matapédia (remplaçant Matapédia) et Gaspé.
- Mégantic-Compton est modifiée et reprend le nom de Mégantic, disparu depuis 1973.
- Montmagny-L'Islet et une partie de Kamouraska-Témiscouata sont regroupés pour former Côte-du-Sud.
- Nicolet-Yamaska disparaît et Nicolet-Bécancour apparaît.
- Prévost se réduit à la ville de Saint-Jérôme.
- Création de Repentigny (détachée de L'Assomption).
- Rivière-du-Loup et une partie de Kamouraska-Témiscouata sont regroupées pour former Rivière-du-Loup–Témiscouata.
- Création de Sainte-Rose (détachée de Fabre et de Vimont).
- Création de Sanguinet (détaché de Châteauguay, de La Prairie et de Huntingdon).
- Shefford est divisée entre Brome-Missisquoi et la nouvelle circonscription de Granby.
- Vanier devient Vanier-Les Rivières.

Un bon nombre de circonscriptions sont également modifiées, souvent pour simplifier le découpage ou refléter les limites de municipalités et d'arrondissements.

Annexe 2 : Description des variables :

Variable	Type de variable		description
Y_PLQ	Variable à expliquer	Quantitative continue	Pourcentage de vote pour le parti libéral
Y_PQ	Variable à expliquer	Quantitative continue	Pourcentage de vote pour le parti québécois
Y_ADQ	Variable à expliquer	Quantitative continue	Pourcentage de vote pour l'ADQ/CAQ
R_parti	Variable explicative	Quantitative continue	Pourcentage des votes pour le parti des élections précédentes.
P_parti	Variable explicative	Quantitative continue	Pourcentage des intentions de votes (sondage précédant le jour des élections de deux semaines)
Candidat	Variable explicative	Binaire	Re-candidature du député du parti qui a gagné les élections précédentes.
Masculin	Variable explicative	Quantitative continue	Fraction des hommes dans la circonscription
Féminin	Variable explicative	Quantitative continue	Fraction des femmes dans la circonscription
age_moyen	Variable explicative	Quantitative continue	Âge moyen dans la circonscription
langue 1	Variable explicative	Quantitative continue	Fraction de la population ayant une connaissance de l'Anglais seulement
langue 2	Variable explicative	Quantitative continue	Fraction de la population ayant une connaissance du français seulement
langue 3	Variable explicative	Quantitative continue	Fraction de la population ayant une connaissance de l'anglais et français
langue 4	Variable explicative	Quantitative continue	Fraction de la population n'ayant aucune connaissance de ses deux langues
non_immigrant	Variable explicative	Quantitative continue	Fraction des non immigrants dans la circonscription
Immigrant	Variable explicative	Quantitative continue	Fraction des immigrants dans la circonscription
tx_chomage	Variable explicative	Quantitative continue	Taux de chômage dans la circonscription
rvn_median	Variable explicative	Quantitative continue	Revenu médian dans la circonscription
D_0	Variable explicative	Quantitative continue	Pourcentage des personnes n'ayant aucun certificat, diplôme ou grade
D_1	Variable explicative	Quantitative continue	Pourcentage des personnes ayant un Diplôme d'études secondaires ou l'équivalent
D_2	Variable explicative	Quantitative continue	Pourcentage des personnes ayant un certificat ou diplôme d'apprenti ou d'une école de métiers
D_3	Variable explicative	Quantitative continue	Pourcentage des personnes ayant un certificat ou diplôme d'un collège, d'un cégep ou d'un autre établissement d'enseignement non universitaire
D_4	Variable explicative	Quantitative continue	Pourcentage des personnes ayant un certificat ou diplôme universitaire en dessous du baccalauréat
D_5	Variable explicative	Quantitative continue	Pourcentage des personnes ayant un certificat ou grade universitaire
Erreur_PLQ	Variable explicative	Quantitative continue	Pourcentage de l'erreur de sondage* des élections précédentes pour le parti libéral
Erreur_PQ	Variable explicative	Quantitative continue	Pourcentage de l'erreur de sondage* des élections précédentes pour le parti québécois
Erreur_ADQ	Variable explicative	Quantitative continue	Pourcentage de l'erreur de sondage* des élections précédentes pour l'ADQ/CAQ

Annexe 3 : Liste des interactions

Modèle 3: Prédiction du pourcentage de vote					
inter1	R_PLQ* PIB	inter32	R_PQ* PIB	inter63	R_ADQ* PIB
inter2	R_PLQ* tx_chomage	inter33	R_PQ* tx_chomage	inter64	R_ADQ* tx_chomage
inter3	R_PLQ* candidat	inter34	R_PQ* candidat	inter65	R_ADQ* candidat
inter4	R_PLQ* immigrant	inter35	R_PQ* immigrant	inter66	R_ADQ* immigrant
inter5	R_PLQ* langue2	inter36	R_PQ* langue2	inter67	R_ADQ* langue2
inter6	R_PLQ* langue3	inter37	R_PQ* langue3	inter68	R_ADQ* langue3
inter7	R_PLQ* langue4	inter38	R_PQ* langue4	inter69	R_ADQ* langue4
inter8	R_PLQ* D_1	inter39	R_PQ* D_1	inter70	R_ADQ* D_1
inter9	R_PLQ* D_2	inter40	R_PQ* D_2	inter71	R_ADQ* D_2
inter10	R_PLQ* D_3	inter41	R_PQ* D_3	inter72	R_ADQ* D_3
inter11	R_PLQ* D_4	inter42	R_PQ* D_4	inter73	R_ADQ* D_4
inter12	R_PLQ* D_5	inter43	R_PQ* D_5	inter74	R_ADQ* D_5
inter13	R_PLQ* masculin	inter44	R_PQ* masculin	inter75	R_ADQ* masculin
inter14	R_PLQ* age_moyen	inter45	R_PQ* age_moyen	inter76	R_ADQ* age_moyen
inter15	R_PLQ* rvn_median	inter46	R_PQ* rvn_median	inter77	R_ADQ* rvn_median
inter16	P_PLQ_NEW*PIB	inter47	P_PQ_NEW*PIB	inter78	P_ADQ_NEW*PIB
inter17	P_PLQ_NEW* tx_chomage	inter48	P_PQ_NEW* tx_chomage	inter79	P_ADQ_NEW* tx_chomage
inter18	P_PLQ_NEW* candidat	inter49	P_PQ_NEW* candidat	inter80	P_ADQ_NEW* candidat
inter19	P_PLQ_NEW* immigrant	inter50	P_PQ_NEW* immigrant	inter81	P_ADQ_NEW* immigrant
inter20	P_PLQ_NEW* langue2	inter51	P_PQ_NEW* langue2	inter82	P_ADQ_NEW* langue2
inter21	P_PLQ_NEW* langue3	inter52	P_PQ_NEW* langue3	inter83	P_ADQ_NEW* langue3
inter22	P_PLQ_NEW* langue4	inter53	P_PQ_NEW* langue4	inter84	P_ADQ_NEW* langue4
inter23	P_PLQ_NEW* D_1	inter54	P_PQ_NEW* D_1	inter85	P_ADQ_NEW* D_1
inter24	P_PLQ_NEW* D_2	inter55	P_PQ_NEW* D_2	inter86	P_ADQ_NEW* D_2
inter25	P_PLQ_NEW* D_3	inter56	P_PQ_NEW* D_3	inter87	P_ADQ_NEW* D_3
inter26	P_PLQ_NEW* D_4	inter57	P_PQ_NEW* D_4	inter88	P_ADQ_NEW* D_4
inter27	P_PLQ_NEW* D_5	inter58	P_PQ_NEW* D_5	inter89	P_ADQ_NEW* D_5
inter28	P_PLQ_NEW* masculin	inter59	P_PQ_NEW* masculin	inter90	P_ADQ_NEW* masculin
inter29	P_PLQ_NEW* age_moyen	inter60	P_PQ_NEW* age_moyen	inter91	P_ADQ_NEW* age_moyen
inter30	P_PLQ_NEW* rvn_median	inter61	P_PQ_NEW* rvn_median	inter92	P_ADQ_NEW* rvn_median
inter31	R_PLQ*P_PLQ_NEW	inter62	R_PQ*P_PQ_NEW	inter93	R_ADQ*P_ADQ_NEW

Modèle 4: Pr�vision de l'erreur de sondages			
inter1	Erreur_PLQ*PIB	inter25	Erreur_PQ* D_2
inter2	Erreur_PLQ* tx_chomage	inter26	Erreur_PQ* D_3
inter3	Erreur_PLQ* candidat	inter27	Erreur_PQ* D_4
inter4	Erreur_PLQ* immigrant	inter28	Erreur_PQ* D_5
inter5	Erreur_PLQ* langue2	inter29	Erreur_PQ* masculin
inter6	Erreur_PLQ* langue3	inter30	Erreur_PQ* age_moyen
inter7	Erreur_PLQ* langue4	inter31	Erreur_PQ* r�n_median
inter8	Erreur_PLQ* D_1	inter32	R_PQ*Erreur_PQ
inter9	Erreur_PLQ* D_2	inter33	Erreur_ADQ*PIB
inter10	Erreur_PLQ* D_3	inter34	Erreur_ADQ* tx_chomage
inter11	Erreur_PLQ* D_4	inter35	Erreur_ADQ* candidat
inter12	Erreur_PLQ* D_5	inter36	Erreur_ADQ* immigrant
inter13	Erreur_PLQ* masculin	inter37	Erreur_ADQ* langue2
inter14	Erreur_PLQ* age_moyen	inter38	Erreur_ADQ* langue3
inter15	Erreur_PLQ* r�n_median	inter39	Erreur_ADQ* langue4
inter16	R_PLQ*Erreur_PLQ	inter40	Erreur_ADQ* D_1
inter17	Erreur_PQ*PIB	inter41	Erreur_ADQ* D_2
inter18	Erreur_PQ* tx_chomage	inter42	Erreur_ADQ* D_3
inter19	Erreur_PQ* candidat	inter43	Erreur_ADQ* D_4
inter20	Erreur_PQ* immigrant	inter44	Erreur_ADQ* D_5
inter21	Erreur_PQ* langue2	inter45	Erreur_ADQ* masculin
inter22	Erreur_PQ* langue3	inter46	Erreur_ADQ* age_moyen
inter23	Erreur_PQ* langue4	inter47	Erreur_ADQ* r�n_median
inter24	Erreur_PQ* D_1	inter48	R_ADQ*Erreur_ADQ

Annexe 4: tableaux des variables significatives:
Modèle de prévision du pourcentage de vote:
Scénario 1 :

2008			2012		
PLQ	PQ	ADQ	PLQ	PQ	CAQ
R_plq	R_pq	R_adq	R_PLQ	R_PQ	R_ADQ
P_PLQ_new	P_pq_new	P_adq_new	P_PLQ_new	P_pq_new	D_3
D_2	D_4	candidat	langue2	langue3	D_5
D_4	masculin	masculin	D_4		
age_moyen	PIB	PIB	D_5		
rvn_median			masculin		

Scénario 2 :

2008			2012		
PLQ	PQ	ADQ	PLQ	PQ	CAQ
inter7	masculin	candidat	R_PLQ	P_pq_new	PIB
inter11	inter32	inter71	langue2	PIB	D_3
inter14	inter42	inter80	masculin	inter42	inter63
inter22	inter62	inter84	inter1	inter43	inter80
inter24	masculin	inter89	inter11	inter44	inter86
inter30			inter12		PIB
inter7					D_3
inter11					

Modèle de prévision de l'erreur de sondage:
Scénario 1 :

2008			2012		
PLQ	PQ	ADQ	PLQ	PQ	CAQ
erreur_plq	erreur_pq	erreur_adq	erreur_PLQ	erreur_PQ	erreur_ADQ
D_1	D_1	PIB	D_4	candidat	R_ADQ
age_moyen	age_moyen	age_moyen		immigrant	
				masculin	

Scénario 2 :

2008			2012		
PLQ	PQ	ADQ	PLQ	PQ	CAQ
tx_chomage	PIB	PIB	D_2	immigrant	PIB
D_1	langue2	D_4	D_4	inter17	inter33
inter1	D_5	age_moyen	inter1	inter24	inter35
inter11	age_moyen	inter41	inter15	inter29	inter45
inter12	rvn_median	inter43		inter32	inter48
		inter44			

Annexe 5: Codes SAS

Modèle 3: prévisions du pourcentage de votes :

Élections 2008 :

Scénario 1 :

```
ods rtf;

/*plq*/
proc glmselect data=Aout_mem.predict2007_125;
model y_PLQ = r_plq p_plq_new PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5 masculin age_moyen
rvn_median/selection=stepwise(select=AIC choose=aic) ;
score data=Aout_mem.predict2008_125 out=Aout_mem.predict2008rsltPLQ
p=P_Y_PLQ;
run;

data Aout_mem.predict2008rsltPLQ; set Aout_mem.predict2008rsltPLQ; keep
circonscription P_Y_PLQ P_PLQ_new;
run;

/*pq*/
proc glmselect data=Aout_mem.predict2007_125;
model y_PQ = r_pq p_pq_new PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5 masculin age_moyen rvn_median/
selection=stepwise(select=AIC choose=aic) ;
score data=Aout_mem.predict2008_125 out=Aout_mem.predict2008rsltPQ
p=P_Y_PQ;
run;

data Aout_mem.predict2008rsltPQ; set Aout_mem.predict2008rsltPQ; keep
circonscription P_Y_PQ P_PQ_new;
run;

/*ADQ*/
proc glmselect data=Aout_mem.predict2007_125;
model y_ADQ = r_adq p_adq_new PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5
masculin age_moyen rvn_median
/ selection=stepwise(select=AIC choose=aic) ;
score data=Aout_mem.predict2008_125 out=Aout_mem.predict2008rsltADQ
p=P_Y_ADQ;
run;

data Aout_mem.predict2008rsltADQ; set Aout_mem.predict2008rsltADQ; keep
circonscription P_Y_ADQ P_ADQ_new;
run;

data Aout_mem.predict2008rslt;
merge Aout_mem.predict2008rsltPLQ Aout_mem.predict2008rsltPQ
Aout_mem.predict2008rsltADQ ;
run;
/*calcul du resultat predit (sondage-erreur)= R_P_PLQ... et le coder
(winner)*/

data Aout_mem.final2; set Aout_mem.predict2008rslt;
M=max(P_Y_PLQ, P_Y_PQ, P_Y_ADQ);
if P_Y_PLQ=M then winner=1; else if P_Y_PQ=M then winner=2; else if
P_Y_ADQ=M then winner=3; run;
```

```

/*comparaison du résultat prédit avec le résultat réel */
data Aout_mem.resultat;
merge Aout_mem.final2 Aout_mem.resultat2008_125; keep circonscription
winner parti_elu ;
run;

/*création du tableau de vérité*/
data Aout_mem.resultat2;
set Aout_mem.resultat;
    if winner=1 and parti_elu=1 then TV=11; else if winner=1 and
parti_elu=2 then TV=12 ;
else if winner=1 and parti_elu=3 then TV=13; else if winner=1 and
parti_elu=4 then TV=14;
else if winner=2 and parti_elu=1 then TV=21; else if winner=2 and
parti_elu=2 then TV=22;
else if winner=2 and parti_elu=3 then TV=23; else if winner=2 and
parti_elu=4 then TV=24;
else if winner=3 and parti_elu=1 then TV=31; else if winner=3 and
parti_elu=2 then TV=32;
else if winner=3 and parti_elu=3 then TV=33; else if winner=3 and
parti_elu=4 then TV=34;
run;
proc freq data=Aout_mem.resultat2;
tables TV ;
run;

/* calcul de l'erreur de prévision
data Aout_mem.residu08S;
merge Aout_mem.final2 Aout_mem.rslt08_parti_125; keep circonscription
P_Y_PLQ P_Y_PQ P_Y_ADQ R_PLQ R_PQ R_ADQ;
run;
data Aout_mem.residu08S;
set Aout_mem.residu08S;

residu_PLQ_08=(P_Y_PLQ-R_PLQ)*(P_Y_PLQ-R_PLQ);
residu_PQ_08=(P_Y_PQ-R_PQ)*(P_Y_PQ-R_PQ);
residu_ADQ_08=(P_Y_ADQ-R_ADQ)*(P_Y_ADQ-R_ADQ);

run;
proc means data=Aout_mem.residu08S;
run;
ods rtf close;

```

Scénario 2 :

```

ods rtf;

/*creation des interactions dans le fichier 2007 et 2008*/
data aout_mem.predict2007_125;
set aout_mem.predict2007_125;

inter1=R_PLQ* PIB; inter2=R_PLQ* tx_chomage; inter3 =R_PLQ* candidat;
inter4=R_PLQ* immigrant; inter5 =R_PLQ* langue2; inter6=R_PLQ* langue3;
inter7=R_PLQ*langue4;inter8=R_PLQ* D_1; inter9=R_PLQ*D_2;inter10=R_PLQ*D_3;
inter11=R_PLQ* D_4; inter12=R_PLQ* D_5; inter13=R_PLQ*masculin;
inter14=R_PLQ* age_moyen ;
inter15=R_PLQ* rvn_median;inter16=P_PLQ_NEW*PIB;inter17=P_PLQ_NEW*
tx_chomage;
inter18=P_PLQ_NEW* candidat;inter19=P_PLQ_NEW* immigrant;inter20=P_PLQ_NEW*
langue2;

```



```

inter21=P_PLQ_NEW* langue3;inter22=P_PLQ_NEW* langue4;inter23=P_PLQ_NEW*
D_1;
inter24=P_PLQ_NEW* D_2;inter25=P_PLQ_NEW* D_3; inter26=P_PLQ_NEW* D_4;
inter27=P_PLQ_NEW* D_5; inter28=P_PLQ_NEW* masculin ;inter29=P_PLQ_NEW*
age_moyen;
inter30=P_PLQ_NEW* rvn_median;inter31=R_PLQ*P_PLQ_NEW;inter32=R_PQ*PIB ;
inter33=R_PQ*tx_chomage;inter34 =R_PQ*candidat; inter35=R_PQ*immigrant ;
inter36=R_PQ*langue2; inter37=R_PQ*langue3; inter38=R_PQ*langue4;
inter39=R_PQ*D_1; inter40=R_PQ*D_2; inter41=R_PQ*D_3; inter42=R_PQ*D_4;
inter43=R_PQ*D_5; inter44=R_PQ*masculin;inter45=R_PQ*age_moyen; inter46
=R_PQ*rvn_median;
inter47=P_PQ_NEW*PIB;inter48=P_PQ_NEW*tx_chomage;inter49=P_PQ_NEW*candidat;
inter50=P_PQ_NEW*immigrant; inter51=P_PQ_NEW*langue2
;inter52=P_PQ_NEW*langue3;
inter53=P_PQ_NEW*langue4;inter54=P_PQ_NEW*D_1;inter55=P_PQ_NEW*D_2;inter56=
P_PQ_NEW*D_3;
inter57=P_PQ_NEW*D_4;inter58=P_PQ_NEW*D_5;inter59=P_PQ_NEW*masculin;
inter60=P_PQ_NEW*age_moyen; inter61=P_PQ_NEW*rvn_median;
inter62=R_PQ*P_PQ_NEW; inter63=R_ADQ*PIB; inter64=R_ADQ*tx_chomage;
inter65=R_ADQ*candidat; inter66=R_ADQ*immigrant; inter67=R_ADQ*langue2;
inter68=R_ADQ*langue3; inter69=R_ADQ*langue4 ; inter70=R_ADQ*D_1 ;
inter71=R_ADQ*D_2 ; inter72=R_ADQ*D_3 ; inter73=R_ADQ*D_4 ;
inter74=R_ADQ*D_5 ; inter75=R_ADQ*masculin ; inter76=R_ADQ*age_moyen ;
inter77=R_ADQ*rvn_median ; inter78=_ADQ_NEW*PIB;
inter79=P_ADQ_NEW*tx_chomage;
inter80=P_ADQ_NEW*candidat;
inter81=P_ADQ_NEW*immigrant;inter82=P_ADQ_NEW*langue2;
inter83=P_ADQ_NEW*langue3;inter84=P_ADQ_NEW*langue4 ;inter85=P_ADQ_NEW*D_1
;
inter86=P_ADQ_NEW*D_2;inter87=P_ADQ_NEW*D_3;inter88=P_ADQ_NEW*D_4 ;
inter89=P_ADQ_NEW*D_5;inter90=P_ADQ_NEW*masculin;inter91=P_ADQ_NEW*age_moye
n ;
inter92=P_ADQ_NEW*rvn_median;inter93=R_ADQ*P_ADQ_NEW;

run;

data aout_mem.predict2008_125;
set aout_mem.predict2008_125;

inter1=R_PLQ* PIB; inter2=R_PLQ* tx_chomage; inter3 =R_PLQ* candidat;
inter4=R_PLQ* immigrant; inter5 =R_PLQ* langue2; inter6=R_PLQ* langue3;
inter7=R_PLQ*langue4;inter8=R_PLQ* D_1; inter9=R_PLQ*D_2;inter10=R_PLQ*D_3;
inter11=R_PLQ* D_4; inter12=R_PLQ* D_5; inter13=R_PLQ*masculin;
inter14=R_PLQ* age_moyen ;
inter15=R_PLQ*
rvn_median;inter16=P_PLQ_NEW*PIB;inter17=P_PLQ_NEW*
tx_chomage;
inter18=P_PLQ_NEW* candidat;inter19=P_PLQ_NEW* immigrant;inter20=P_PLQ_NEW*
langue2;
inter21=P_PLQ_NEW* langue3;inter22=P_PLQ_NEW* langue4;inter23=P_PLQ_NEW*
D_1;
inter24=P_PLQ_NEW* D_2;inter25=P_PLQ_NEW* D_3; inter26=P_PLQ_NEW* D_4;
inter27=P_PLQ_NEW* D_5; inter28=P_PLQ_NEW* masculin ;inter29=P_PLQ_NEW*
age_moyen;
inter30=P_PLQ_NEW* rvn_median;inter31=R_PLQ*P_PLQ_NEW;inter32=R_PQ*PIB ;
inter33=R_PQ*tx_chomage;inter34 =R_PQ*candidat; inter35=R_PQ*immigrant ;
inter36=R_PQ*langue2; inter37=R_PQ*langue3; inter38=R_PQ*langue4;
inter39=R_PQ*D_1; inter40=R_PQ*D_2; inter41=R_PQ*D_3; inter42=R_PQ*D_4;
inter43=R_PQ*D_5; inter44=R_PQ*masculin;inter45=R_PQ*age_moyen; inter46
=R_PQ*rvn_median;
inter47=P_PQ_NEW*PIB;inter48=P_PQ_NEW*tx_chomage;inter49=P_PQ_NEW*candidat;
inter50=P_PQ_NEW*immigrant; inter51=P_PQ_NEW*langue2
;inter52=P_PQ_NEW*langue3;

```

```

inter53=P_PQ_NEW*langue4;inter54=P_PQ_NEW*D_1;inter55=P_PQ_NEW*D_2;inter56=
P_PQ_NEW*D_3;
inter57=P_PQ_NEW*D_4;inter58=P_PQ_NEW*D_5;inter59=P_PQ_NEW*masculin;
inter60=P_PQ_NEW*age_moyen; inter61=P_PQ_NEW*rvn_median;
inter62=R_PQ*P_PQ_NEW; inter63=R_ADQ*PIB; inter64=R_ADQ*tx_chomage;
inter65=R_ADQ*candidat; inter66=R_ADQ*immigrant; inter67=R_ADQ*langue2;
inter68=R_ADQ*langue3; inter69=R_ADQ*langue4 ; inter70=R_ADQ*D_1 ;
inter71=R_ADQ*D_2 ; inter72=R_ADQ*D_3 ; inter73=R_ADQ*D_4 ;
inter74=R_ADQ*D_5 ; inter75=R_ADQ*masculin ; inter76=R_ADQ*age_moyen ;
inter77=R_ADQ*rvn_median ; inter78=_ADQ_NEW*PIB;
inter79=P_ADQ_NEW*tx_chomage;
inter80=P_ADQ_NEW*candidat;
inter81=P_ADQ_NEW*immigrant;inter82=P_ADQ_NEW*langue2;
inter83=P_ADQ_NEW*langue3;inter84=P_ADQ_NEW*langue4 ;inter85=P_ADQ_NEW*D_1
;
inter86=P_ADQ_NEW*D_2;inter87=P_ADQ_NEW*D_3;inter88=P_ADQ_NEW*D_4 ;
inter89=P_ADQ_NEW*D_5;inter90=P_ADQ_NEW*masculin;inter91=P_ADQ_NEW*age_moye
n ;
inter92=P_ADQ_NEW*rvn_median;inter93=R_ADQ*P_ADQ_NEW;
run;

proc glmselect data=aout_mem.predict2007_125 ;
model y_PLQ = r_plq p_plq_new PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5 masculin age_moyen rvn_median inter1-
inter31 / selection=stepwise(select=AIC choose=aic) ;
score data=aout_mem.predict2008_125 out=aout_mem.predictPLQ08_inter
p=P_Y_PLQ r=R_Y_PLQ ;
run;
data aout_mem.predictPLQ08_inter; set aout_mem.predictPLQ08_inter; keep
circonscription P_Y_PLQ P_PLQ_new R_Y_PLQ;
run;

proc glmselect data=aout_mem.predict2007_125 ;
model y_PQ = r_pq p_pq_new PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5 masculin age_moyen rvn_median
inter32-inter62/ selection=stepwise(select=AIC choose=aic) ;
score data=aout_mem.predict2008_125 out=aout_mem.predictPQ08_inter p=P_Y_PQ
r=R_Y_PQ;
run;
data aout_mem.predictPQ08_inter; set aout_mem.predictPQ08_inter; keep
circonscription P_Y_PQ P_PQ_new R_Y_PQ;
run;

proc glmselect data=aout_mem.predict2007_125 ;
model y_adQ = r_adq p_adq_new PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5 masculin age_moyen rvn_median
inter63-inter93 / selection=stepwise(select=AIC choose=aic) ;
score data=aout_mem.predict2008_125 out=aout_mem.predictADQ08_inter
p=P_Y_ADQ r=R_Y_ADQ;
run;

data aout_mem.predictADQ08_inter; set aout_mem.predictADQ08_inter; keep
circonscription P_Y_ADQ P_ADQ_new R_Y_ADQ;
run;

data aout_mem.predict08rslt;
merge aout_mem.predictPLQ08_inter aout_mem.predictPQ08_inter
aout_mem.predictADQ08_inter ;
run;

/*calcul du resultat predit (sondage-erreur)= R_P_PLQ... et le coder
(winner)*/

```

```

data aout_mem.final2; set aout_mem.predict08rslt;
M=max(P_Y_plq, P_Y_pq, P_Y_adq);
if P_Y_plq=M then winner=1; else if P_Y_pq=M then winner=2; else if
P_Y_adq=M then winner=3;
run;

/*comparaison du resultat prédit avec le resultat reel 2008*/

data aout_mem.resultat;
merge aout_mem.final2 aout_mem.resultat2008_125; keep circonscription
winner parti_elu ;
run;

/*creation du tableau de verité*/
data aout_mem.resultat2;
set aout_mem.resultat;
if winner=1 and parti_elu=1 then TV=11; else if winner=1 and
parti_elu=2 then TV=12 ;
else if winner=1 and parti_elu=3 then TV=13; else if winner=1 and
parti_elu=4 then TV=14;
else if winner=2 and parti_elu=1 then TV=21; else if winner=2 and
parti_elu=2 then TV=22;
else if winner=2 and parti_elu=3 then TV=23; else if winner=2 and
parti_elu=4 then TV=24;
else if winner=3 and parti_elu=1 then TV=31; else if winner=3 and
parti_elu=2 then TV=32;
else if winner=3 and parti_elu=3 then TV=33; else if winner=3 and
parti_elu=4 then TV=34;

run;

proc freq data=aout_mem.resultat2;
tables TV ;
run;

/*Calcul de l'erreur de prévision
data aout_mem.residu08I;
merge aout_mem.final2 aout_mem.rslt08_parti_125; keep circonscription
P_Y_PLQ P_Y_PQ P_Y_ADQ R_PLQ R_PQ R_ADQ;
run;

data aout_mem.residu08I;
set aout_mem.residu08I;

residu_PLQ_08=(P_Y_PLQ-R_PLQ)*(P_Y_PLQ-R_PLQ);
residu_PQ_08=(P_Y_PQ-R_PQ)*(P_Y_PQ-R_PQ);
residu_ADQ_08=(P_Y_ADQ-R_ADQ)*(P_Y_ADQ-R_ADQ);
run;
proc means data=aout_mem.residu08I;
run;
ods rtf close;

```

Modèle 4: prévisions de l'erreur de sondages :

Élections 2008 :

Scénario 1 :

```
ods rtf;
proc glmselect data=erreurl1.erreur07_125;
model y_PLQ = erreur_PLQ R_PLQ PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5 masculin age_moyen tx_chomage
rvn_median / selection=stepwise(select=AIC choose=aic) ;
score data=erreurl1.erreur08_125 out=erreurl1.erreur08rsltPLQ p=P_Y_PLQ ;
run;

data erreurl1.erreur08rsltPLQ; set erreurl1.erreur08rsltPLQ; keep
circonscription P_Y_PLQ P_PLQ_new; run;

proc glmselect data=erreurl1.erreur07_125 ;
model y_PQ = erreur_PQ R_PQ PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5 masculin age_moyen tx_chomage
rvn_median / selection=stepwise(select=AIC choose=aic) ;
score data=erreurl1.erreur08_125 out=erreurl1.erreur08rsltPQ p=P_Y_PQ ;
run;

data erreurl1.erreur08rsltPQ; set erreurl1.erreur08rsltPQ; keep
circonscription P_Y_PQ P_PQ_NEW; run;
proc glmselect data=erreurl1.erreur07_125 ;
model y_ADQ = erreur_adQ R_ADQ PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5
masculin age_moyen tx_chomage rvn_median
/ selection=stepwise(select=AIC choose=aic) ;
score data=erreurl1.erreur08_125 out=erreurl1.erreur08rsltADQ p=P_Y_ADQ ;
run;

data erreurl1.erreur08rsltADQ; set erreurl1.erreur08rsltADQ; keep
circonscription P_Y_ADQ P_ADQ_NEW; run;

data erreurl1.erreur08rslt;
merge erreurl1.erreur08rsltPLQ erreurl1.erreur08rsltPQ
erreurl1.erreur08rsltADQ ;
run;

/*calcul du resultat predit (sondage-erreur)= R_P_PLQ... et le coder
(winner)*/

data erreurl1.final08S; set erreurl1.erreur08rslt;
R_P_PLQ=P_PLQ_NEW-P_Y_PLQ; R_P_PQ=P_PQ_NEW-P_Y_PQ; R_P_ADQ=P_ADQ_NEW-
P_Y_ADQ;
M=max(R_P_PLQ, R_P_PQ, R_P_ADQ);
if R_P_PLQ=M then winner=1; else if R_P_PQ=M then winner=2; else if
R_P_ADQ=M then winner=3;
run;
/*comparaison du resultat prédit avec le resultat reel 2012*/

data erreurl1.resultat08s;
merge erreurl1.final08s erreurl1.resultat2008_125; keep circonscription
winner parti_elu ;
run;

/*creation du tableau de verité*/
```

```

data erreur11.resultat08s2;
set erreur11.resultat08s;
    if winner=1 and parti_elu=1 then TV=11; else if winner=1 and
parti_elu=2 then TV=12 ;
else if winner=1 and parti_elu=3 then TV=13; else if winner=1 and
parti_elu=4 then TV=14;
else if winner=2 and parti_elu=1 then TV=21; else if winner=2 and
parti_elu=2 then TV=22;
else if winner=2 and parti_elu=3 then TV=23; else if winner=2 and
parti_elu=4 then TV=24;
else if winner=3 and parti_elu=1 then TV=31; else if winner=3 and
parti_elu=2 then TV=32;
else if winner=3 and parti_elu=3 then TV=33; else if winner=3 and
parti_elu=4 then TV=34;

run;

proc freq data=erreur11.resultat08s2;
tables TV ;
run;
ods rtf close;

/*Calcul de l'erreur de prévision

data erreur11.residu08S;
merge erreur11.final08S erreur11.rslt08_parti_125; keep circonscription
R_P_PLQ R_P_PQ R_P_ADQ R_PLQ R_PQ R_ADQ;
run;

data erreur11.residu08S;
set erreur11.residu08S;

residu_PLQ_08=(R_P_PLQ-R_PLQ)*(R_P_PLQ-R_PLQ);
residu_PQ_08=(R_P_PQ-R_PQ)*(R_P_PQ-R_PQ);
residu_ADQ_08=(R_P_ADQ-R_ADQ)*(R_P_ADQ-R_ADQ);

run;

proc means data=erreur11.residu08S;
run;
ods rtf close;

```

Scénario 2 :

```

ods rtf;

/*erreur 2008*/
data erreur11.erreur07_125;          set          erreur11.erreur07_125;
inter1=Erreur_PLQ*PIB;
inter2=Erreur_PLQ*          tx_chomage;          inter3=Erreur_PLQ*          candidat;
inter4=Erreur_PLQ* immigrant;
inter5=Erreur_PLQ* langue2; inter6=Erreur_PLQ* langue3; inter7=Erreur_PLQ*
langue4;
inter8=Erreur_PLQ* D_1; inter9=Erreur_PLQ* D_2; inter10=Erreur_PLQ* D_3;
inter11=Erreur_PLQ* D_4; inter12=Erreur_PLQ* D_5; inter13=Erreur_PLQ*
masculin ;
inter14=Erreur_PLQ*          age_moyen;          inter15=Erreur_PLQ*
rvn_median;inter16=R_PLQ*Erreur_PLQ;
inter17=Erreur_PQ*PIB; inter18=Erreur_PQ* tx_chomage;inter19=Erreur_PQ*
candidat ;
inter20=Erreur_PQ* immigrant;inter21=Erreur_PQ*langue2;inter22=Erreur_PQ*
langue3 ;
inter23=Erreur_PQ* langue4;inter24=Erreur_PQ* D_1 ;inter25=Erreur_PQ* D_2
;
inter26=Erreur_PQ* D_3;inter27=Erreur_PQ* D_4;inter28=Erreur_PQ* D_5 ;
inter29=Erreur_PQ*masculin;inter30=Erreur_PQ*age_moyen;inter31=Erreur_PQ*
rvn_median;inter32=R_PQ*Erreur_PQ;inter33=Erreur_ADQ*PIB;
inter34=Erreur_ADQ*          tx_chomage
;inter35=Erreur_ADQ*candidat;inter36=Erreur_ADQ*
immigrant;inter37=Erreur_ADQ* langue2 ;inter38=Erreur_ADQ* langue3 ;
inter39=Erreur_ADQ* langue4;inter40=Erreur_ADQ* D_1 ;inter41=Erreur_ADQ*
D_2 ;
inter42=Erreur_ADQ* D_3 ;inter43=Erreur_ADQ* D_4 ;inter44=Erreur_ADQ*
D_5 ;
inter45=Erreur_ADQ*masculin;inter46=Erreur_ADQ*age_moyen;inter47=Erreur_ADQ
* rvn_median;inter48=R_ADQ*Erreur_ADQ;

run;

data erreur11.erreur08_125;
set erreur11.erreur08_125;

inter1=Erreur_PLQ*PIB;
inter2=Erreur_PLQ*          tx_chomage;          inter3=Erreur_PLQ*          candidat;
inter4=Erreur_PLQ* immigrant;
inter5=Erreur_PLQ* langue2; inter6=Erreur_PLQ* langue3; inter7=Erreur_PLQ*
langue4;
inter8=Erreur_PLQ* D_1; inter9=Erreur_PLQ* D_2; inter10=Erreur_PLQ* D_3;
inter11=Erreur_PLQ* D_4; inter12=Erreur_PLQ* D_5; inter13=Erreur_PLQ*
masculin ;
inter14=Erreur_PLQ*          age_moyen;          inter15=Erreur_PLQ*
rvn_median;inter16=R_PLQ*Erreur_PLQ;
inter17=Erreur_PQ*PIB; inter18=Erreur_PQ* tx_chomage;inter19=Erreur_PQ*
candidat ;
inter20=Erreur_PQ* immigrant;inter21=Erreur_PQ*langue2;inter22=Erreur_PQ*
langue3 ;
inter23=Erreur_PQ* langue4;inter24=Erreur_PQ* D_1 ;inter25=Erreur_PQ* D_2
;
inter26=Erreur_PQ* D_3;inter27=Erreur_PQ* D_4;inter28=Erreur_PQ* D_5 ;
inter29=Erreur_PQ*masculin;inter30=Erreur_PQ*age_moyen;inter31=Erreur_PQ*
rvn_median;inter32=R_PQ*Erreur_PQ;inter33=Erreur_ADQ*PIB;

```

```

inter34=Erreur_ADQ*                                tx_chomage
;inter35=Erreur_ADQ*candidat;inter36=Erreur_ADQ*
immigrant;inter37=Erreur_ADQ* langue2 ;inter38=Erreur_ADQ* langue3 ;
inter39=Erreur_ADQ* langue4;inter40=Erreur_ADQ* D_1 ;inter41=Erreur_ADQ*
D_2 ;
inter42=Erreur_ADQ* D_3 ;inter43=Erreur_ADQ* D_4 ;inter44=Erreur_ADQ*
D_5 ;
inter45=Erreur_ADQ*masculin;inter46=Erreur_ADQ*age_moyen;inter47=Erreur_ADQ
* rvn_median;inter48=R_ADQ*Erreur_ADQ;

run;

proc glmselect data=erreurl1.erreur07_125 ;
model y_PLQ = erreur_PLQ R_PLQ PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5 masculin age_moyen rvn_median
inter1-inter16 / selection=stepwise(select=AIC choose=aic) ;
score data=erreurl1.erreur08_125 out=erreurl1.erreurPLQ08_inter p=P_Y_PLQ ;
run;

data erreurl1.erreurPLQ08_inter; set erreurl1.erreurPLQ08_inter; keep
circonscription P_Y_PLQ P_PLQ_new;b
run;

proc glmselect data=erreurl1.erreur07_125 ;
model y_PQ = erreur_PQ R_PQ PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5 masculin age_moyen rvn_median
inter17-inter32 / selection=stepwise(select=AIC choose=aic) ;
score data=erreurl1.erreur08_125 out=erreurl1.erreurPQ08_inter p=P_Y_PQ ;
run;

data erreurl1.erreurPQ08_inter; set erreurl1.erreurPQ08_inter; keep
circonscription P_Y_PQ P_PQ_new;
run;

proc glmselect data=erreurl1.erreur07_125 ;
model y_adQ = erreur_adQ R_ADQ PIB tx_chomage candidat immigrant langue2
langue3 langue4 D_1 D_2 D_3 D_4 D_5 masculin age_moyen rvn_median
inter33-inter48 / selection=stepwise(select=AIC choose=aic) ;
score data=erreurl1.erreur08_125 out=erreurl1.erreurADQ08_inter p=P_Y_ADQ ;
run;

data erreurl1.erreurADQ08_inter; set erreurl1.erreurADQ08_inter; keep
circonscription P_Y_ADQ P_ADQ_new;
run;

data erreurl1.erreur08_inter;
merge erreurl1.erreurPLQ08_inter erreurl1.erreurPQ08_inter
erreurl1.erreurADQ08_inter; keep circonscription P_PLQ_NEW P_PQ_NEW
P_ADQ_NEW P_Y_PLQ
P_Y_PQ P_Y_ADQ;
run;
/*calcul du resultat predit (sondage-erreur)= R_P_PLQ... et le coder
(winner)*/

data erreurl1.final_inter; set erreurl1.erreur08_inter;
R_P_PLQ=P_PLQ_NEW-P_Y_PLQ; R_P_PQ=P_PQ_NEW-P_Y_PQ; R_P_ADQ=P_ADQ_NEW-
P_Y_ADQ;
M=max(R_P_PLQ, R_P_PQ, R_P_ADQ);
if R_P_PLQ=M then winner=1; else if R_P_PQ=M then winner=2; else if
R_P_ADQ=M then winner=3;
run;
/*comparaison du resultat prédit avec le resultat reel 2012*/

data erreurl1.resultat_erreur;

```

```

merge      erreur11.final_inter      erreur11.resultat2008_125;keep
circonscription winner parti_elu ;
run;

/*creation du tableau de verité*/
data erreur11.resultat2_erreur;
set erreur11.resultat_erreur;
    if winner=1 and parti_elu=1 then TV=11; else if winner=1 and
parti_elu=2 then TV=12 ;
else if winner=1 and parti_elu=3 then TV=13; else if winner=1 and
parti_elu=4 then TV=14;
else if winner=2 and parti_elu=1 then TV=21; else if winner=2 and
parti_elu=2 then TV=22;
else if winner=2 and parti_elu=3 then TV=23; else if winner=2 and
parti_elu=4 then TV=24;
else if winner=3 and parti_elu=1 then TV=31; else if winner=3 and
parti_elu=2 then TV=32;
else if winner=3 and parti_elu=3 then TV=33; else if winner=3 and
parti_elu=4 then TV=34;

run;

proc freq data=erreur11.resultat2_erreur;
tables TV ;
run;

/*calcul de l'erreur de prévision*/

data erreur11.residu08I;
merge erreur11.final_inter erreur11.rslt08_parti_125; keep circonscription
R_P_PLQ R_P_PQ R_P_ADQ R_PLQ R_PQ R_ADQ;
run;

data erreur11.residu08I;
set erreur11.residu08I;

residu_PLQ_08=(R_P_PLQ-R_PLQ)*(R_P_PLQ-R_PLQ);
residu_PQ_08=(R_P_PQ-R_PQ)*(R_P_PQ-R_PQ);
residu_ADQ_08=(R_P_ADQ-R_ADQ)*(R_P_ADQ-R_ADQ);

run;

proc means data=erreur11.residu08I;
run;

ods rtf close;

```


**Annexe 6a: Sorties SAS modèle 3.
Élections 2008 :**

Scénario 1 :

PLQ

Root MSE	0.07158
Dependent Mean	0.34007
R-Square	0.7704
Adj R-Sq	0.7587
AIC	-525.44057
AICC	-524.19919
SBC	-632.64237

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.277969	0.153648	-1.81
R_plq	1	0.588106	0.053652	10.96
P_PLQ_new	1	0.496599	0.069026	7.19
D_2	1	-0.705893	0.274396	-2.57
D_4	1	-3.304178	0.995835	-3.32
age_moyen	1	0.007337	0.003234	2.27
rvn_median	1	0.00000595	0.00000217	2.73

PQ :

Root MSE	0.06272
Dependent Mean	0.28584
R-Square	0.6970
Adj R-Sq	0.6842
AIC	-559.40660
AICC	-558.44934
SBC	-669.43672

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.585189	0.233611	-2.50
R_pq	1	0.575779	0.055332	10.41
P_pq_new	1	0.525585	0.093696	5.61
PIB	1	0.182051	0.123116	1.48
D_4	1	1.197569	0.488209	2.45
masculin	1	0.932049	0.446685	2.09

ADQ :

Root MSE	0.06482
Dependent Mean	0.29508
R-Square	0.7601
Adj R-Sq	0.7500
AIC	-551.16764
AICC	-550.21037
SBC	-661.19775

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.276936	0.202568	-1.37
R_adq	1	0.794528	0.075410	10.54
P_adq_new	1	0.574652	0.085762	6.70
PIB	1	-0.231146	0.124604	-1.86
candidat	1	0.025183	0.013777	1.83
masculin	1	0.571217	0.407820	1.40

Scénario 2 :

PLQ :

Root MSE	0.06828
Dependent Mean	0.34007
R-Square	0.7911
Adj R-Sq	0.7804
AIC	-537.23625
AICC	-535.99487
SBC	-644.43805

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.101995	0.022843	-4.46
inter7	1	-7.327729	3.060790	-2.39
inter11	1	-5.484352	1.605613	-3.42
inter14	1	0.023226	0.002248	10.33
inter22	1	11.958852	3.937322	3.04
inter24	1	-1.152817	0.553166	-2.08
inter30	1	0.00002235	0.00000405	5.51

PQ :

Root MSE	0.06450
Dependent Mean	0.28584
R-Square	0.6769
Adj R-Sq	0.6661
AIC	-553.38335
AICC	-552.67149
SBC	-666.24178

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.288307	0.216121	-1.33
masculin	1	0.740306	0.434080	1.71
inter32	1	0.678494	0.315377	2.15
inter42	1	3.216146	1.281398	2.51
inter62	1	1.437924	0.188969	7.61

ADQ :

Root MSE	0.06138
Dependent Mean	0.29508
R-Square	0.7849
Adj R-Sq	0.7759
AIC	-564.83418
AICC	-563.87692
SBC	-674.86430

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	0.080336	0.018340	4.38
candidat	1	-0.079793	0.027280	-2.92
inter71	1	5.164954	0.411178	12.56
inter80	1	0.427024	0.105048	4.07
inter84	1	4.115858	1.671136	2.46
inter89	1	0.991330	0.275641	3.60

Élections 2012 :

Scénario 1 :

PLQ :

Root MSE	0.07734
Dependent Mean	0.43985
R-Square	0.7434
Adj R-Sq	0.7287
AIC	-452.55812
AICC	-451.16007
SBC	-547.52863

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	0.872061	0.303617	2.87
R_PLQ	1	0.749241	0.061147	12.25
P_PLQ_new	1	0.187412	0.075884	2.47
langue2	1	0.073421	0.049487	1.48
D_4	1	-4.191380	1.134611	-3.69
D_5	1	0.536673	0.147669	3.63
masculin	1	-1.436955	0.578446	-2.48

PQ :

Root MSE	0.07587
Dependent Mean	0.34969
R-Square	0.6634
Adj R-Sq	0.6541
AIC	-459.71970
AICC	-459.15366
SBC	-562.84570

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.014393	0.031874	-0.45
R_PQ	1	0.733444	0.074301	9.87
P_pq_new	1	0.346030	0.085098	4.07
langue3	1	0.087970	0.041911	2.10

CAQ :

Root MSE	0.05279
Dependent Mean	0.14292
R-Square	0.7254
Adj R-Sq	0.7178
AIC	-540.95018
AICC	-540.38414
SBC	-644.07618

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.101485	0.037071	-2.74
R_ADQ	1	0.651938	0.039535	16.49
D_3	1	0.483265	0.229619	2.10
D_5	1	-0.110783	0.051559	-2.15

Scénario 2 :PLQ :

Root MSE	0.07378
Dependent Mean	0.43985
R-Square	0.7687
Adj R-Sq	0.7531
AIC	-462.19311
AICC	-460.42841
SBC	-554.44512

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	1.083452	0.269579	4.02
R_PLQ	1	0.754340	0.175183	4.31
langue2	1	0.059317	0.042149	1.41
masculin	1	-1.909800	0.554878	-3.44
inter1	1	1.582891	1.155249	1.37
inter11	1	-13.862593	2.803507	-4.94
inter12	1	1.724306	0.407239	4.23
inter31	1	0.622489	0.178130	3.49

PQ:

Root MSE	0.06524
Dependent Mean	0.34969
R-Square	0.7649
Adj R-Sq	0.7442
AIC	-487.91867
AICC	-485.27867
SBC	-574.73368

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	0.062853	0.067008	0.94
P_pq_new	1	-0.565490	0.295034	-1.92
PIB	1	-3.740966	1.330121	-2.81
inter42	1	-29.01302	7.006428	-4.14
inter43	1	4.326198	0.916949	4.72
inter44	1	3.863639	0.574922	6.72
inter47	1	8.209665	3.393797	2.42
inter57	1	33.659400	6.277909	5.36
inter58	1	-4.422679	0.838395	-5.28
inter62	1	-1.286716	0.620391	-2.07

CAQ:

Root MSE	0.04802
Dependent Mean	0.14292
R-Square	0.7770
Adj R-Sq	0.7665
AIC	-560.25121
AICC	-559.17428
SBC	-657.94021

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	0.034656	0.036269	0.96
PIB	1	-2.601582	0.342515	-7.60
D_3	1	0.404225	0.212018	1.91
inter63	1	10.001521	0.742545	13.47
inter80	1	-0.151184	0.098798	-1.53
inter86	1	2.114756	0.564728	3.74

Annexe 6b: Sorties SAS modèle 4.

Élections 2008 :

Scénario 1 :

PLQ :

Root MSE	0.08626
Dependent Mean	0.02283
R-Square	0.4493
Adj R-Sq	0.4357
AIC	-481.65477
AICC	-481.15057
SBC	-597.34151

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	0.542792	0.148061	3.67
erreur_plq	1	0.519350	0.054064	9.61
D_1	1	-0.747331	0.329352	-2.27
age_moyen	1	-0.007299	0.003227	-2.26

PQ :

Root MSE	0.06352
Dependent Mean	0.00575
R-Square	0.5889
Adj R-Sq	0.5787
AIC	-558.14744
AICC	-557.64323
SBC	-673.83418

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.018813	0.106803	-0.18
R_pq	1	-0.639764	0.050289	-12.72
D_1	1	0.408766	0.244195	1.67
age_moyen	1	0.003777	0.002380	1.59

ADQ :

Root MSE	0.06410
Dependent Mean	-0.04493
R-Square	0.5410
Adj R-Sq	0.5296
AIC	-555.90544
AICC	-555.40124
SBC	-671.59218

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.282474	0.092734	-3.05
erreur_adq	1	0.700876	0.060362	11.61
PIB	1	0.288207	0.121630	2.37
age_moyen	1	0.005965	0.002357	2.53

Scénario 2 :PLQ :

Root MSE	0.08118
Dependent Mean	0.02283
R-Square	0.5244
Adj R-Sq	0.5002
AIC	-493.96406
AICC	-492.72268
SBC	-601.16586

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	0.454533	0.095511	4.76
tx_chomage	1	-0.490692	0.291004	-1.69
D_1	1	-1.485998	0.375613	-3.96
inter1	1	1.830401	1.123648	1.63
inter11	1	-18.44439	5.234673	-3.52
inter12	1	2.973157	0.694914	4.28
inter14	1	0.021824	0.004937	4.42

PQ :

Root MSE	0.05952
Dependent Mean	0.00575
R-Square	0.6541
Adj R-Sq	0.6302
AIC	-569.71777
AICC	-567.78794
SBC	-671.26294

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.346302	0.124990	-2.77
PIB	1	-0.205430	0.128844	-1.59
langue2	1	-0.069273	0.037181	-1.86
D_5	1	-0.214363	0.085757	-2.50
age_moyen	1	0.007433	0.002680	2.77
rvn_median	1	0.00000357	0.00000159	2.25
inter17	1	3.148730	0.908404	3.47
inter28	1	1.007218	0.513687	1.96
inter30	1	0.008532	0.002576	3.31

ADQ :

Root MSE	0.06135
Dependent Mean	-0.04493
R-Square	0.5900
Adj R-Sq	0.5691
AIC	-564.00512
AICC	-562.76374
SBC	-671.20692

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.279424	0.090941	-3.07
PIB	1	0.269460	0.121673	2.21
D_4	1	-0.858167	0.456127	-1.88
age_moyen	1	0.006856	0.002283	3.00
inter41	1	6.439607	1.063671	6.05
inter43	1	-26.488878	7.969253	-3.32
inter44	1	6.086186	1.569256	3.88

Élections 2012 :

Scénario 1 :

PLQ :

Root MSE	0.06177
Dependent Mean	0.02158
R-Square	0.7604
Adj R-Sq	0.7560
AIC	-506.74317
AICC	-506.36934
SBC	-612.58767

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.034037	0.021700	-1.57
erreur_PLQ	1	0.934710	0.050815	18.39
D_4	1	0.773409	0.438965	1.76

PQ :

Root MSE	0.07002
Dependent Mean	-0.00281
R-Square	0.5934
Adj R-Sq	0.5782
AIC	-476.73755
AICC	-475.93755
SBC	-577.14505

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	0.398420	0.248915	1.60
erreur_PQ	1	0.846722	0.070538	12.00
candidat	1	-0.031973	0.019312	-1.66
immigrant	1	-0.146434	0.054767	-2.67
masculin	1	-0.739458	0.499396	-1.48

CAQ :

Root MSE	0.05176
Dependent Mean	-0.03203
R-Square	0.5890
Adj R-Sq	0.5814
AIC	-546.34179
AICC	-545.96795
SBC	-652.18629

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	0.015181	0.013495	1.12
erreur_ADQ	1	0.605715	0.070867	8.55
R_ADQ	1	-0.082142	0.049948	-1.64

Scénario 2 :

PLQ :

Root MSE	0.06054
Dependent Mean	0.02158
R-Square	0.7740
Adj R-Sq	0.7656
AIC	-509.29622
AICC	-508.49622
SBC	-609.70373

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.128421	0.070291	-1.83
D_2	1	0.331683	0.227448	1.46
D_4	1	1.641480	0.788826	2.08
inter1	1	4.624930	3.021331	1.53
inter15	1	0.000029731	0.000006133	4.85

PQ :

Root MSE	0.06793
Dependent Mean	-0.00281
R-Square	0.6209
Adj R-Sq	0.6030
AIC	-482.56248
AICC	-481.48556
SBC	-580.25149

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.005041	0.008922	-0.56
immigrant	1	-0.102608	0.048676	-2.11
inter17	1	4.768736	3.377760	1.41
inter24	1	-4.387284	2.233330	-1.96
inter29	1	3.867206	1.006685	3.84
inter32	1	-1.033991	0.502191	-2.06

CAQ :

Root MSE	0.04083
Dependent Mean	-0.03203
R-Square	0.7513
Adj R-Sq	0.7396
AIC	-596.60709
AICC	-595.53017
SBC	-694.29610

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-0.032593	0.012403	-2.63
PIB	1	0.695862	0.227593	3.06
inter33	1	10.643181	2.360394	4.51
inter35	1	0.153781	0.101766	1.51
inter45	1	-0.932456	0.320834	-2.91
inter48	1	1.281990	0.266670	4.81