

# Les méthodes de mesure de l'importance des critères de satisfaction

## *Application dans le domaine du service automobile*

Pierre WINDAL  
Directeur de la société PWC

et  
Pierre DESMET  
Professeur à l'Université Paris IX-Dauphine et à l'ESSEC

Dans le secteur automobile, la mesure régulière de la satisfaction est aussi importante en terme de service qu'en terme de produit. Dans une perspective de produit, elle est indispensable pour éclairer les choix pour un véhicule futur ou pour améliorer un véhicule existant. Dans une perspective de service elle est également indispensable, plus fréquemment encore, aux différentes occasions de contact avec les clients (vente d'un véhicule, livraison ou après-vente); la satisfaction du service conduit en effet, tant à des comportements de répétition d'achat (figure 1) qu'à la diffusion d'une communication non maîtrisée par le bouche-à-oreille.

Les données issues de ces enquêtes de satisfaction de service sont utilisées dans deux cadres :

- *Dans une optique d'analyse*, pour comprendre le processus conduisant à la satisfaction, ou à l'insatisfaction, et isoler les facteurs les plus importants sur lesquels il est nécessaire d'agir.
- *Dans une optique tactique*, pour suivre la performance d'un centre de profit (une concession) et identifier les actions de progrès permettant de l'améliorer. Cette dimension a d'autant plus d'importance que la performance est souvent retenue comme indicateur pour les éléments variables de la rémunération.

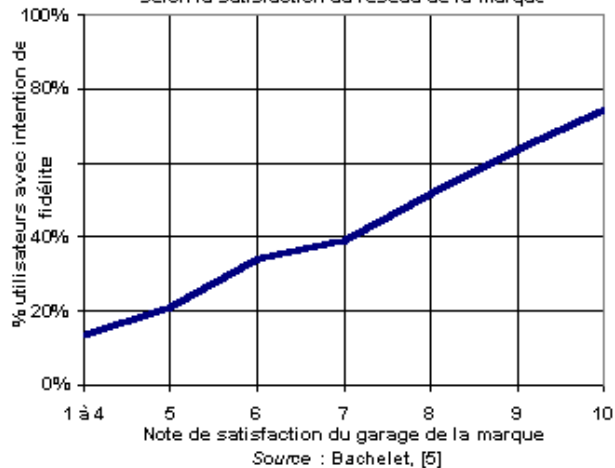
Pour l'analyse et les recommandations, le modèle le plus fréquemment utilisé considère que la satisfaction globale provient d'une satisfaction sur différents critères, pondérée par leur importance. Or, pour estimer l'importance des critères, les études qualitatives ont montré que le questionnement direct des interviewés conduisait à des biais importants en faveur des facteurs rationnels (par exemple le prix de la réparation) et en défaveur des facteurs plus subjectifs (par exemple l'amabilité du personnel). Dans la pratique actuelle, les poids sont donc obtenus par une approche statistique basée sur les corrélations entre la satisfaction globale et les satisfactions par critères.

Mais les praticiens se heurtent alors souvent à des difficultés importantes. En effet les critères, pour être opérationnels, doivent couvrir les différentes facettes de la satisfaction et être suffisamment détaillés ; ils sont donc assez nombreux, par exemple 20 à 50, parfois plus. Ils sont alors toujours corrélés entre eux, ce qui conduit à

des résultats qui n'ont pas les caractéristiques attendues sur le plan de la vraisemblance, de la stabilité ou de la robustesse et qui sont donc inexploitable. Le praticien se trouve donc confronté à une difficulté concrète récurrente (les enquêtes sont périodiques) et mal résolue par l'ensemble des techniques actuellement disponibles.

Cet article a donc pour objectif de relancer le débat sur les problèmes pratiques que pose la mesure de l'importance de critères corrélés dans le cadre des études de satisfaction. D'où les questions abordées : dans un contexte managérial, qu'est-ce qu'une bonne mesure de satisfaction et comment la mettre en œuvre ? Quelles sont les méthodes actuellement utilisées, leurs spécificités et leurs mérites respectifs selon différents contextes ?

Figure 1 : intention de fidélité à la marque selon la satisfaction du réseau de la marque



Pour préciser concrètement les performances de ces méthodes, l'article comparera les résultats de trois méthodes appliquées à une enquête sur la satisfaction du service à l'achat et à livraison d'un véhicule.

La problématique serait naturellement la même pour une enquête de satisfaction concernant le véhicule, ou pour tout autre service ou produit dont la satisfaction comporterait des critères nombreux (en pratique plus de 6 ou 8).

## Les difficultés de mesure

Le cadre général de la mesure de la satisfaction ayant déjà été traité dans un numéro spécial de la revue [1], il ne sera pas repris ici.

Dans le domaine des enquêtes de satisfaction sur le service automobile, le croisement de deux dimensions, chronologique et fonctionnelle, génère souvent la colinéarité. On s'intéressera, par exemple, à la performance du personnel de la concession en différentes occasions (achat, livraison, entretien ...) sur de multiples critères (*disponibilité, compétence, organisation ...*). On suivra également certains critères « sensibles » pour le constructeur automobile (*aspect des locaux, variété des modèles exposés*), mais pas toujours très fortement liés à la satisfaction globale du client.

Cette cohabitation forcée d'items explicatifs redondants et d'items spécifiques ayant une influence faible sur le phénomène étudié, conduit généralement à l'exclusion des critères les moins explicatifs de chaque dimension. Prenons l'exemple de l'amabilité du personnel de la concession. Elle est mesurée à diverses occasions (achat, livraison,...) et sous diverses formes (écoute, disponibilité,...). Le client est sensible à la notion globale d'*amabilité du personnel*, qu'elle se manifeste au moment de l'achat ou de la livraison de sa voiture. Le constructeur, lui, s'intéresse au détail : en cas de problèmes d'amabilité du personnel, il veut savoir qui est concerné, du personnel affecté à l'achat ou à la livraison. Si les deux se valent, le client tendra à noter de la même manière l'aspect « *amabilité du personnel à l'achat* » et l'aspect « *amabilité du personnel à la livraison* ». En régression multiple classique, seul l'un de ces deux items sera doté d'un coefficient significatif. Peut-on en conclure que l'un est important et l'autre pas ? Non. On peut en revanche conclure que la connaissance du second n'apporte rien à l'explication du phénomène étudié au-delà de la connaissance du premier.

Pour mieux comprendre les origines de cette multicollinéarité, il est nécessaire de s'intéresser aux données traitées. La colinéarité peut, en effet, venir d'une information insuffisante (faible nombre d'observations), être structurellement liée au problème étudié (relation taille de la concession, amabilité), provenir d'une redondance volontaire pour obtenir une meilleure mesure d'une dimension par plusieurs items ou même résulter du questionnaire par lui-même.

Dans la pratique actuelle, les enquêtes de satisfaction dans l'automobile correspondent à deux objectifs : les *enquêtes inter-constructeurs* (ECS) comparent la performance des réseaux des différentes marques alors que les *enquêtes par concession* comparent la performance des concessions d'une marque. Les premières sont annuelles, postales, avec un effectif de 200 à 500 répondants par marque et par pays, soit environ 10000 par pays. Les secondes sont réalisées par téléphone, typiquement environ 2000 clients, de manière tournante sur les concessions.

Les effectifs des études concessions étudiées ici peuvent paraître importants, mais se révèlent pourtant souvent contraignants lorsqu'il s'agit de travailler sur des segments de clientèle. La rotation effectuée sur les concessions permet d'alléger le coût de l'étude tout en bénéficiant d'informations périodiques sur chaque concession. Elle conduit cependant à une assez forte variation des informations et à une instabilité des matrices de corrélations des variables d'une vague sur l'autre. Ceci se traduit par le fait que l'importance estimée des critères peut varier significativement entre deux collectes, ce qui est difficilement acceptable.

Le questionnaire en lui-même contient une trentaine de critères sur la satisfaction globale, par dimension et par item, ainsi que sur les différents facteurs pouvant l'influencer. Ces variables sont souvent fortement corrélées, sans même parler de la multi-collinéarité : il n'est pas rare d'observer des corrélations directes de 0.5 entre les dimensions ou de 0.8 entre les variables d'une même dimension.

Lorsqu'elle est connue, une structure hiérarchique (dimension, items) peut être imposée. Les questions sur les items de chaque dimension ne sont alors posées qu'en fonction de questions filtres, générales, sur la dimension. Dans de nombreux cas, les champs des items sont donc vides et l'information manquante, ce qui peut entraîner des difficultés lorsque les méthodes ne permettent pas un traitement facile des données manquantes. Inversement, lorsque les données manquantes sont peu nombreuses, l'analyse peut ne pas être facilitée par la prise en compte de données collectées de manière plus ou moins « forcée » par l'enquêteur.

La rotation des questions n'est pas une pratique habituelle tant parce que le questionnaire suit une logique de « visite d'achat » allant de l'accueil à la reprise du véhicule, que parce que cela complexifie le processus de l'étude et risque donc d'introduire des erreurs. Il est donc intéressant de souligner que, parfois, l'ordre de l'importance des critères correspond à celui de la proximité de l'échelle d'évaluation globale, confirmant ainsi l'effet d'ordre attendu.

Sur le plan de l'ajustement global, la qualité des résultats obtenus n'est que moyenne, la corrélation entre les critères et la satisfaction globale étant rarement forte au niveau individuel. Typiquement, avec des coefficients de détermination ( $R^2$ ) situés entre 0.5 et 0.6, les modèles actuels n'expliquent qu'un peu plus de la moitié des variations de la satisfaction pour les individus interrogés.

Enfin, les estimateurs des poids ne sont pas toujours cohérents avec les attentes des décideurs ou avec les résultats des études qualitatives préalables, tant au niveau de chaque coefficient (signe négatif, valeur nulle) qu'au niveau de la hiérarchie des critères selon leur importance (critère important ayant un poids faible et

inversement).

poids des critères. Cette approche, correspondant à une analyse bayésienne, a pourtant l'inconvénient d'introduire une inertie dans le modèle, le coefficient utilisé pouvant mettre plusieurs périodes avant de s'adapter aux nouveaux poids créés par une modification du marché.

.....

### Le cahier des charges

Lorsqu'un modèle doit être utilisé dans un cadre normatif, pour l'aide à la décision, il est connu depuis longtemps que les caractéristiques du modèle aux yeux de son utilisateur (le décideur) jouent un rôle au moins aussi important que celui de la qualité de l'ajustement statistique [3]. Le modèle n'est vraiment utilisable, et utilisé, que s'il a les caractéristiques suivantes :

La *simplicité* signifie que le modèle doit pouvoir être mis à jour et modifié rapidement. La *robustesse* indique la capacité du modèle à fournir des réponses cohérentes même lorsque les données qui lui sont fournies sont très différentes des données ayant été utilisées pour le calibrer. La *complétude* permet au décideur de retrouver dans le modèle des résultats cohérents avec ceux d'autres études, en particulier qualitatives, et avec ceux issus de l'expérience. Enfin, la *compréhension* exige que le modèle soit facilement compris par le décideur.

Pour chaque variable, le décideur souhaite obtenir une sensibilité reliant la variation de la variable et sa conséquence en terme de variation de la satisfaction. Les deux questions à résoudre sont d'une part, les variables à retenir, et d'autre part, le poids à leur attribuer pour reconstituer la satisfaction globale du répondant.

*Le choix des variables à retenir est délicat.* En effet, pour éviter l'oubli d'un critère important, c'est la redondance qui est choisie avec l'intégration de nombreux items issus de réunions de groupes. A côté des variables influençant la satisfaction, sont ainsi pris en compte des variables mesurant la satisfaction aux différentes étapes du service.

La structuration préalable des variables, sous forme d'un modèle structurel ou d'une arborescence, permet de bénéficier d'un modèle théorique. Cet arbre est construit par expertise ou par recours à une étude qualitative structurée de type analyse en chaînes moyens-fins.

*La pondération des déterminants de la satisfaction* peut être obtenue par inférence statistique à un niveau individuel ou agrégé :

- à un niveau individuel, l'inférence statistique par une méthode d'analyse conjointe réclame un nombre important de données et le recours à des poids « self expliqués » fournis par le répondant peut être nécessaire comme dans les modèles hybrides.
- à un niveau agrégé, l'évaluation statistique des poids a pour conséquence une instabilité des estimateurs du fait des variations aléatoires de chaque échantillon. Une consolidation des poids est donc préférable, chaque vague d'enquêtes permettant de mettre à jour, par une prise en compte partielle, le

Dans la suite de l'article, nous considérons que ces questions ont été traitées et nous nous concentrons sur l'évaluation statistique, pour un échantillon donné et sans connaissance a priori, des poids d'un ensemble déterminé de critères.

.....

### Hypothèses du modèle et colinéarité

Le modèle reliant les composantes et la satisfaction est supposé unique et linéaire, hypothèses lourdes qui ne seront cependant pas remises en cause dans cet article.

- *l'unicité du modèle* signifie que tous les clients partagent le même système de pondération des critères. La remise en cause de cette hypothèse passe par une segmentation a priori (sur des critères déterminés au préalable), a posteriori (sur base de coefficients d'importance des critères obtenus au niveau individuel) ou par une recherche simultanée de modèles explicatifs par segments (régression avec classes latentes). L'unicité du modèle, alors que le marché est composé des segments ayant des processus de choix différents, est une des raisons qui peuvent expliquer la médiocre qualité de l'ajustement global.

- *la linéarité du modèle* a pour conséquence l'absence d'effet d'interaction entre les variables : le coefficient reliant une variable à la satisfaction ne dépend pas des valeurs des autres variables. Sauf à introduire des variables d'interaction, il n'est donc pas possible de représenter une hiérarchie de choix complexe dans laquelle la prise en compte ou non d'une variable dépendrait du résultat de l'évaluation préalable d'un autre critère. Cette linéarité permet cependant d'avoir un modèle simple et s'adaptant assez bien à des situations différentes.

Dans un modèle linéaire où, par commodité, les variables sont centrées et réduites, le rôle de la colinéarité est clairement mis en évidence par l'équation normale de la méthode des moindres carrés :

.....

$$(eq.1) \quad r_i = b_i + \sum_{j \neq i} b_j r_{ij} \text{ où :}$$

- $r_i$  dénote le coefficient de corrélation simple entre la variable  $i$  et le critère à expliquer.
- $b_i$  est le coefficient de régression de la variable  $i$ .
- $r_{ij}$  est le coefficient de corrélation entre les variables  $i$  et  $j$ .

Le coefficient de corrélation simple entre la variable  $i$  et le critère à expliquer est égal à la somme de deux effets : (1) l'importance propre de la variable :  $b_i$  et (2) l'importance qu'elle partage avec les autres critères :  $\sum_{j \neq i} b_j r_{ij}$ .

Il est clair qu'à  $r_i$  fixe, le cumul des variables conduit mécaniquement le coefficient de régression  $b_i$  vers zéro, puis en dessous de zéro. Si l'on dupliquait  $n$  fois la même variable, par exemple, le poids de la variable serait égal à  $r_i/n$ .

Le problème vient de ce que la méthode des moindres carrés a été conçue pour estimer les paramètres d'un modèle où chaque variable explicative compte, et non en gérer le surnombre ou la redondance. Elle s'applique donc bien aux situations où l'objectif consiste à identifier un modèle parcimonieux, et mal, au contraire, aux situations où l'on cherche à donner un poids à chaque variable. Le problème est mal posé au départ, d'où le caractère parfois insatisfaisant de la solution (inversion de signe et coefficient statistiquement nul).

La présence de variables avec erreur de mesure parmi les variables explicatives, la colinéarité de celles-ci, voire même l'inversion de la relation lorsque des effets de halo existent, sont des violations des hypothèses de base de la régression linéaire qui peuvent sérieusement perturber la qualité des résultats. A cela peut s'ajouter la violation de l'hypothèse de normalité des résidus fréquente lorsque les échelles de mesure sont catégorielles, de type Likert, et ne comportent que peu de réponses possibles ou que les réponses se concentrent sur quelques valeurs.

Notamment, la colinéarité peut affecter le signe et la valeur des coefficients, ainsi que leur écart-type : des variables considérées a priori comme importantes héritent d'un poids faible voire négatif ! Ce problème a conduit à l'utilisation de méthodes effectuant des régressions successives pour donner à ces variables l'opportunité d'avoir un poids lorsque les effets des variables auxquelles elles sont corrélées ont été pris en compte.

### Les solutions pratiquées

Elles visent à atténuer, sans totalement la supprimer, la prépondérance des  $r_{ij}$  dans l'équation normale (eq.1). La solution la plus simple consiste à en réduire l'ordre de grandeur, c'est-à-dire à donner plus de poids à l'importance partagée et moins de poids à l'importance propre. C'est ce que font les deux méthodes

habituellement préconisées [3] en présence de forte colinéarité : la régression sur composantes principales et la régression ridge.

#### *La régression sur composantes principales*

Dans la régression, les variables initiales sont remplacées par un sous-ensemble de leurs composantes principales (ACP), après une rotation orthogonale Varimax, en général celles qui sont le plus corrélées à la variable à expliquer. Les coefficients des variables initiales sont ensuite recalculés à partir de ceux des composantes. Il y a donc successivement changement de base et réduction de la dimensionalité du problème.

Pour voir en quoi cette pratique réduit l'importance moyenne des  $r_{ij}$ , rappelons que ceux-ci peuvent être exprimés en termes des valeurs ( $\lambda_k$ ) et vecteurs ( $u_{ik}$ ) propres de la matrice des corrélations :  $r_{ij} = \sum_k \lambda_k u_{ik} u_{jk}$ . Le fait de négliger certaines composantes dans cette somme tend à réduire la valeur moyenne absolue des corrélations  $r_{ij}$ , donc à réduire la variance des coefficients de régression.

Cette méthode est celle utilisée par l'institut J.D. Power pour le calcul de ses indices de satisfaction du service par marque automobile qu'il publie aux E.U. et au R.U. Cependant la contribution à un axe n'a aucun rapport avec la satisfaction globale et la régression sur composantes principales ne donne donc des poids pertinents que pour les principales dimensions (les composantes principales retenues) ; elle n'offre aucune garantie sur le signe des coefficients et tend à diluer l'importance des critères.

#### *La régression ridge*

Elle consiste à réduire la valeur des corrélations par l'ajout d'une constante positive  $\lambda$  à la diagonale de la matrice des corrélations. Si  $R$  dénote cette matrice et  $r$  le vecteur des corrélations simples entre les prédicteurs et la variable à expliquer, les coefficients  $b(\lambda)$  de régression ridge sont égaux à  $(R + \lambda I)^{-1} r$ . Il a été montré qu'il existe une valeur de  $\lambda$  telle que la variance de  $b(\lambda)$  est inférieure à celle de  $b(0)$ , la solution des moindres carrés ordinaires. La valeur moyenne des corrélations  $r_{ij}$  est réduite d'un facteur de  $1/(1+\lambda)$ .

En pratique, cette méthode n'est guère utilisée du fait de ses inconvénients. La régression ridge laisse en effet trop de place à l'arbitraire quant au choix de la constante (évaluation graphique de la stabilité des coefficients). Elle se prête donc mal à la mesure automatique du poids des critères et si elle permet de rétablir les inversions de signe, elle ne le fait qu'à la marge (coefficient voisin de zéro).

Le salut devait venir d'une nouvelle méthode conçue explicitement pour résoudre les problèmes de la colinéarité, la régression PLS.

#### *La régression des moindres carrés partiels (PLS)*

Elle procède par changement de base, comme la régression sur composantes principales, à cette grande

différence près : ce n'est plus la variance de la composante qui est maximisée, mais la corrélation entre chacune des composantes orthogonales et la variable à expliquer. Le nombre de composantes est déterminé selon un test statistique approprié [4]. Les coefficients de régression ne dépendent, à un facteur de normalisation près, que de leur covariance avec la variable à expliquer. D'où un double avantage : la bonne gestion non seulement des valeurs manquantes (calcul deux à deux), mais aussi de la colinéarité (chaque composante successive pousse la solution de  $r_i$  vers  $b_i$ ).

En pratique, lorsque le nombre de prédicteurs est élevé, plusieurs difficultés subsistent : la régression PLS n'offre aucune garantie contre les inversions de signe au-delà de la première composante. Elle gère bien la colinéarité (critères peu nombreux, mais fortement corrélés), mais moins bien la redondance (critères à colinéarité moyenne, mais très nombreux).

#### *Les méthodes empiriques*

Dans l'attente d'une régression PLS avec contrainte, les praticiens ont développé des procédures, certes empiriques, mais dont le mérite est de valider la principale clause du cahier des charges : le respect du signe de la corrélation simple.

Ces solutions empiriques sont sous-optimales au sens des moindres carrés, c'est-à-dire qu'elles produisent des poids intermédiaires entre deux bornes extrêmes :

$$b_i \leq \text{Poids}_i \leq r_i$$

Le poids d'un critère est, au mieux, égal à son coefficient de corrélation simple ( $r_i$ ), au pire, à son coefficient de régression multiple ( $b_i$ ). Selon les stratégies de gestion de la redondance, on se rapprochera plutôt de l'une ou de l'autre de ces deux bornes.

Les méthodes empiriques s'inspirent du principe suivant : s'il n'est pas satisfaisant d'exclure des prédicteurs, il est impossible de les inclure tous, en une seule passe. D'où l'idée de procéder par étapes, en traitant à chaque étape un sous-ensemble de critères, puis en les fusionnant au dernier stade. La difficulté réside dans le choix d'un critère de fusion. Cette approche empirique sera illustrée par deux méthodes au traitement radicalement opposé de la colinéarité : la régression en cascade (RC) et la régression séquentielle pas-à-pas (RSPP). La régression séquentielle pas-à-pas regroupe à chaque régression les variables les moins corrélées entre elles et les plus corrélées à la variable à expliquer, alors que la régression en cascade prend en compte des nœuds composés d'un petit nombre de variables fortement corrélées.

#### *Les méthodes univariées*

Certains praticiens prônent le recours à des méthodes univariées, comme la régression simple, excluant ainsi toute prise en compte des corrélations entre les prédicteurs. Même si le poids n'est pas assimilé à la corrélation simple – ce serait trop simple – on se rapproche cependant de la borne maximale de

l'intervalle. Cette pratique a des avantages certains dans le cadre d'études longitudinales : réduction de l'erreur d'échantillonnage aux seuls  $r_i$  et non plus  $r_{ij}$ , indépendance du poids à l'ajout ou au retrait de critères corrélés, simplicité de calcul, meilleure gestion des valeurs manquantes etc. C'est, toutefois, une solution qui n'exploite pas toute l'information disponible ( $r_{ij}$ ). Entre exploiter tous les  $r_{ij}$  et n'en exploiter aucun, le balancier va trop loin.

La méthode univariée retenue ici, une parmi d'autre, est simple : si  $S$  dénote la satisfaction globale et  $X$  la satisfaction sur un critère donné, on assimile le poids du critère  $X$  à la proportion de clients insatisfaits « globalement » parmi ceux qui sont insatisfaits sur  $X$  :

$$\text{Poids du critère } X = (\text{nombre de clients insatisfaits sur } S \text{ et } X) / (\text{nombre de clients insatisfaits sur } X)$$

Cet indicateur n'exploite que l'information fournie par la clientèle insatisfaite, celle dont on doit se soucier en priorité. C'est une mesure pertinente, mais qui doit être interprétée à la lumière du pourcentage d'insatisfaits pour éviter les problèmes d'une base de calcul trop faible. En pratique, elle s'avère beaucoup plus réactive que les mesures d'association exploitant l'ensemble des informations (satisfaits et insatisfaits), en particulier lorsque le niveau de satisfaction est très élevé.

### **La régression en cascade**

La régression en cascade, présentée la première fois dans un article de Bachelet [5] et s'inspirant des modèles récursifs, comporte deux étapes : (1) la spécification d'un arbre hiérarchique regroupant les items fortement corrélés sur des nœuds à différents niveaux (dimensions, thèmes, sous-thèmes etc.) et (2) la détermination du poids des variables et des « nœuds » ainsi créés, par régression.

D'un point de vue statistique, on néglige les  $r_i$  des nœuds voisins pour le calcul des coefficients de régression. Bien que les variables d'un même nœud soient souvent (mais pas nécessairement) très corrélées, la faible taille des nœuds (*quelques variables*) et la plus ou moins grande égalité des corrélations simples avec la variable dépendante ( $r_i$ )<sup>i</sup> compensent cette forte intercorrélations ( $r_{ij}$ ).

#### *(1) Spécification de l'arbre hiérarchique*

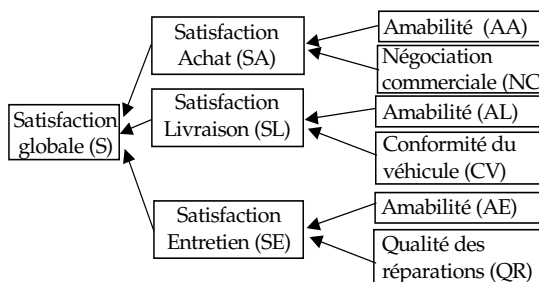
La spécification de l'arbre de régression est laissée au jugement de l'utilisateur, sujet aux contraintes énoncées plus haut sur le lien entre la variable à expliquer, les variables dépendantes intermédiaires et les items explicatifs simples. Le meilleur arbre étant un arbre sans branches, c'est-à-dire le plus proche possible des conditions de la régression multiple ordinaire, on veillera à minimiser le nombre de niveaux hiérarchiques car le poids final d'un item est d'autant plus faible qu'il se trouve en aval de l'arbre<sup>ii</sup>.

La structure de l'arbre reflète souvent celle du questionnaire à la source des données. On préfère, en

général, respecter la logique du questionnement, elle-même fonction de la chronologie ou du sujet. Dans le domaine automobile, par exemple, on parlera d'abord de l'achat puis de la livraison, et on regroupera les items traitant des mêmes dimensions. Les effets d'ordre ne sont donc que partiellement contrôlables par le jeu des rotations d'items, le bénéfice de l'intelligibilité du questionnaire étant jugé supérieur aux avantages de la présentation aléatoire des questions. Ces deux effets – regroupement logique et effet d'ordre – font que l'arbre « sémantique » rejoint souvent l'arbre « statistique », basé sur les seules corrélations.

## (2) Estimation des poids par régression

Dans le cas d'un arbre explicite, les items sont, par exemple, structurés de la façon suivante avec une mesure des variables intermédiaires de satisfaction (figure 2) :



**Figure 2** : Arbre de régression

Cet arbre de régression comprend 10 items, dont la variable à expliquer (*satisfaction globale*) et 4 « nœuds » ou « branches ». On procédera à autant de régressions que de branches. Il faut remonter l'arbre de la droite vers la gauche, les régressions des nœuds en aval devant être effectuées avant celles des nœuds en amont. Seuls les nœuds de même niveau hiérarchique sont interchangeable.

Trois régressions pour le premier niveau : par exemple, une première régression avec *satisfaction à l'achat* pour variable à expliquer par *l'amabilité à l'achat* et la *négociation commerciale*. Une régression au deuxième niveau pour expliquer la *satisfaction globale* par la *satisfaction à l'achat*, la *satisfaction à la livraison* et la *satisfaction à l'entretien*.

Dans le cas d'un arbre implicite, un nœud intermédiaire, par exemple la *satisfaction à l'achat*, ne correspond pas à une variable dépendante explicite. La variable intermédiaire est alors remplacée par la variable à expliquer - *la satisfaction globale* - pour le calcul du poids des items à l'intérieur du nœud, et, ensuite, remplacée par la satisfaction globale reconstituée pour la régression au deuxième niveau<sup>iii</sup>.

Les variables sont toutes centrées-réduites, y compris les variables reconstituées. A chaque nœud, le poids « local » d'un item correspond à son coefficient de régression standardisé, divisé par la somme des coefficients de régression standardisés de tous les items du nœud, de sorte que la somme des poids des items d'un nœud donné soit toujours égale à 1. Le poids

« final » d'un item est égal au produit des poids successifs : son poids local et celui des branches dont il fait partie. Dans l'exemple, le poids de l'item *amabilité à l'achat* serait égal à  $(a_1 \cdot b_1)$ ,  $b_1$  correspondant au coefficient de régression standardisé et normé de la *satisfaction à l'achat* (ou de la variable reconstituée *S*) dans la régression de la *satisfaction globale*.

La régression en cascade est particulièrement adaptée aux situations dans lesquelles il existe une hiérarchie naturelle des critères, allant du plus général au plus détaillé. Dans l'exemple ci-dessus, la *satisfaction à l'achat* englobe *l'amabilité du personnel à l'achat* et la *négociation commerciale*. L'item servant de variable dépendante au nœud doit être plus corrélée à la variable à expliquer que chacun des items constitutifs du nœud, et la corrélation entre la variable dépendante du nœud et les items du nœud doit naturellement être plus élevée qu'aucune autre corrélation. Sinon, ce serait cette autre variable qu'il faudrait prendre comme variable dépendante du nœud.

## La régression séquentielle pas-à-pas

Ici l'ensemble des items est considéré comme un échantillon de sous-ensembles indépendants et, dans une certaine mesure, redondants, chacun pouvant expliquer avec plus ou moins de précision la satisfaction globale. Cette perspective reflète la diversité des instruments utilisés par les constructeurs pour mesurer le même concept (la satisfaction du client) et l'évolution de l'ensemble dans le temps : certains items sont éliminés et d'autres ajoutés, d'autres encore sont spécifiques à un marché.

Le problème du calcul des poids consiste alors à identifier le nombre et la composition des sous-ensembles d'items, le poids des items à l'intérieur de ces groupes et le poids respectif de chaque groupe. Pour résoudre ce problème, la régression séquentielle pas-à-pas procède comme suit (cf. Windal [6]) :

(1) Jusqu'à épuisement des items ou non significativité statistique de tous les items restants, itérer sur la séquence suivante ;

- Effectuer une régression pas-à-pas sur les items restants (au départ : la totalité des items).
- Eliminer les items dotés du "bon" signe et d'un coefficient significatif au sens du "t" de Student. Par "bon" signe, on entend un signe identique à celui de la corrélation simple entre l'item et la variable à expliquer, ou un signe imposé au départ.

Le nombre de prédicteurs et le  $R^2$  de la régression diminuent donc à chaque séquence.

(2) Après la dernière séquence, toute variable restante est entrée seule dans une régression, quels que soient son signe et son degré de significativité.

(3) Le poids final d'un item est égal au produit du  $R^2$  de la régression à laquelle il appartient et de son coefficient

de régression<sup>iv</sup> standardisé et normé à 1 à l'intérieur de sa régression (NB : ici, c'est le coefficient maximal que l'on norme à 1). Le poids du critère le plus explicatif est donc égal au  $R^2$  de sa régression.

Cette procédure présente plusieurs avantages :

- Les poids des variables indépendantes seront toujours significatifs<sup>v</sup> dans leur régression et, par construction, dotés du bon signe, c'est-à-dire celui de leur corrélation simple.
- Les variables non encore retenues entrent en compétition avec toutes les autres variables à toutes les étapes, même la première. Au contraire de la régression en cascade, chaque variable est donc confrontée, au moins une fois, à toutes les autres.
- Deux variables fortement corrélées auront un poids du même ordre de grandeur. Contrairement à ce qui se passe en régression multiple, l'une n'"écrase" pas l'autre. Cette propriété essentielle de la méthode vient de ce que ces variables entreront dans des régressions distinctes.
- La procédure est automatique et donne un résultat unique : pas d'arbre à spécifier, comme en régression en cascade, ni de coefficient atténuateur de colinéarité à choisir, comme en régression ridge.

Le seul point contestable de cette méthode tient à la fixation arbitraire du poids de chaque séquence, égal par définition au  $R^2$  de la séquence, alors que l'information apportée par toutes les séquences au-delà de la première est nulle au plan statistique. Géométriquement, cela revient à prendre comme poids « final » pour l'item de la séquence  $i$  la projection du poids qu'il a dans sa régression sur la droite formant un angle de  $\cos^{-1}(r_{i1})$  avec l'abscisse,  $r_{i1}$  dénotant le coefficient de corrélation entre les deux séquences (figure 3).

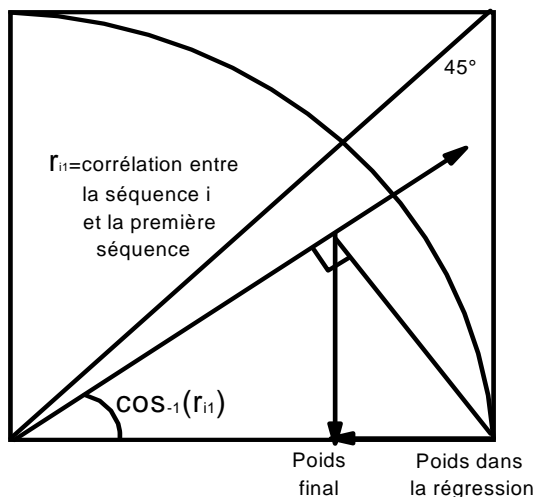


Figure 3 : Calcul du poids final d'un item de la séquence  $i$

### Application au service automobile

Les résultats des trois méthodes de régression, PLS, en

cascade (RC) et séquentielle pas-à-pas (RSPP), vont être comparés entre eux ainsi qu'à la pondération univariée. Le logiciel utilisé est MENSOR<sup>®</sup> de la société Quadrature.

### Les données

Les données sont extraites d'un baromètre de satisfaction à l'égard de l'achat et de la livraison d'un véhicule automobile. Les clients sont interrogés par téléphone très peu de temps après la livraison de leur voiture sur la base d'un questionnaire d'une dizaine de minutes.

Le questionnaire respecte la chronologie du processus d'achat, en invitant le client, en une trentaine de questions, à se remémorer et à évaluer son expérience de l'achat et de la livraison de sa voiture. La question de satisfaction globale est posée en début de questionnaire, pour favoriser la spontanéité de la réponse et neutraliser l'effet d'ordre (inévitabile si la question est posée plus tard).

La batterie d'items comprend une question de satisfaction globale et vingt cinq questions de satisfaction à l'égard de trois grandes dimensions : les locaux, l'achat et la livraison. Ces trois dimensions sont également jugées, d'où un total de 28 notes de satisfaction « détaillée ». La satisfaction est mesurée sur une échelle en quatre points, de « tout-à-fait satisfait » à « tout-à-fait insatisfait ».

### Formalisation de l'arbre

La construction de l'arbre est une étape d'autant plus délicate qu'il n'existe pas de règles garantissant un arbre unique. En outre, des modifications mineures de l'arbre peuvent dégrader fortement la significativité statistique de certains coefficients de régression, sans altérer sensiblement le pourcentage de variance expliquée. On privilégiera donc la validité statistique des coefficients à la précision de l'ajustement ( $R^2$ ).

La régression en cascade prend explicitement en compte la nature hiérarchique de cette batterie d'items et, ici, suit la logique du questionnaire, aux subdivisions près<sup>vi</sup>. Les trois dimensions sont traitées comme des variables dépendantes, via l'« arbre de régression » présenté dans la figure 4.

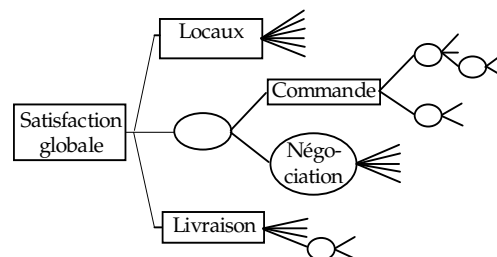


Figure 4 : Arbre de régression

Dans cet arbre de régression, les rectangles correspondent à des variables dépendantes observées et les cercles à des variables dépendantes latentes pour lesquelles la satisfaction globale ou le critère de niveau

supérieur servent de variables instrumentales. L'introduction d'une dimension  *négociation*, suggérée par l'analyse en composantes principales des 25 items, améliore la qualité de l'ajustement. Les autres subdivisions, strictement empiriques, servent à faire en sorte que les coefficients de régression soient tous significatifs.

Pour la régression PLS, le test de validité croisée conduit à retenir deux composantes. Pour la régression séquentielle pas-à-pas, les 25 critères sont traités en trois séquences. Ces deux méthodes de régression (PLS, RSPP) sont automatiques et conduisent toujours aux mêmes résultats. Il faut, toutefois, prendre soin d'exclure de la liste des variables explicatives les variables correspondant aux trois dimensions intermédiaires (locaux, commande et livraison). Ne pas le faire conduirait à donner un poids excessif à ces trois dimensions et, pour cause de colinéarité, à minorer le poids des items les plus importants à l'intérieur de chaque dimension<sup>vii</sup>.

#### *Echantillon de travail et échantillon test*

De la base initiale des 3000 clients, deux sous-échantillons (A et B) de 2000 personnes ont été tirés avec remise pour effectuer une validation croisée : les coefficients de l'un ont été utilisés pour calculer les mesures de performances de l'autre et vice-versa. Cette procédure d'échantillonnage avec remise procure aux deux sous-échantillons une base commune d'environ une observation sur deux, minimisant ainsi le risque de différences significatives en terme de moyennes et de variances. C'est, naturellement, une situation plus favorable que celle à laquelle les décideurs doivent en pratique faire face.

#### *Mesures de performances, de stabilité et de dispersion*

En matière de hiérarchies de critères de satisfaction, qualité de l'ajustement, robustesse et différenciation sont les trois caractéristiques les plus attendues des poids.

La qualité de l'ajustement correspond à la qualité de la reconstitution de la satisfaction globale. Elle est mesurée par le pourcentage de la variance de la satisfaction globale expliquée par les items ( $R^2$ ) et par l'erreur quadratique moyenne rapportée à la moyenne (RMSE). Elle est étudiée sur l'échantillon de travail et, de manière croisée, sur l'échantillon test.

La robustesse est l'aptitude de la méthode à fournir des résultats proches aux variations aléatoires près, et se traduit par la plus ou moins grande stabilité des poids d'un échantillon à l'autre. Ce critère peut correspondre à deux réalités : « *ça bouge éventuellement beaucoup, mais dans le même sens* » (corrélation) et « *ça bouge peu, mais éventuellement dans tous les sens* » (différence de niveau). Selon que l'on est statisticien ou décideur, on sera plus sensible à la première ou à la seconde. La robustesse est donc mesurée par deux indicateurs : le coefficient de corrélation entre les poids des échantillons de travail et de test, et la proportion de poids de l'échantillon test supérieurs de 30%, en valeur absolue,

aux poids de l'échantillon de travail. La robustesse de la méthode se manifeste également à travers la bonne tenue de sa précision en validation croisée (stabilité du  $R^2$ ).

Quant à la dispersion des poids, elle doit être ni trop forte, ni trop faible. A précision égale ( $R^2$ ), on préférera une hiérarchie équilibrée aux hiérarchies plate (peu de différence entre les poids) ou dominante (un poids élevé et de nombreux poids très faibles). L'écart-type des poids de l'échantillon de travail fournit une mesure simple de cette dispersion.

## **Comparaison des résultats**

### *Qualité de l'ajustement et capacité prédictive*

Les trois méthodes de régression sont d'une précision équivalente mais la qualité de l'ajustement n'est que moyenne, puisque le  $R^2$  se situe entre 0.46 et 0.59 selon la méthode et l'échantillon. Ce n'est donc pas un critère déterminant dans le choix de l'une ou l'autre de ces méthodes.

< insérer ici le tableau 1 >

Si la précision de la régression en cascade est légèrement supérieure à celles des deux autres méthodes, c'est surtout grâce à la prise en compte d'un plus grand nombre de variables (les trois dimensions intermédiaires). En revanche, à nombre de variables identiques, elle est en général un peu moins précise que les deux autres, car certaines variables n'interviennent pas directement dans l'équation de régression finale. Cet handicap est le prix à payer pour sa capacité à gérer les données de type hiérarchique.

### *Validité convergente*

La validité convergente exprime le degré de recouvrement des hiérarchies extraites par les différentes méthodes des échantillons de travail et de test. Elle est mesurée, ici, par le coefficient de corrélation entre les poids de chaque paire de méthodes, échantillon par échantillon. Plus la validité convergente est élevée, moins le choix de la méthode porte à conséquence.

< insérer ici le tableau 2 >

Les régressions PLS et RSPP produisent des hiérarchies de critères comparables ( $r=0.8$ ). La validité convergente est du même ordre de grandeur que la robustesse. Le choix de l'une par rapport à l'autre peut reposer sur des considérations non statistiques, telles la simplicité de calcul ou la disponibilité du logiciel.

Les deux autres méthodes conduisent à des hiérarchies non dénuées de mérite, mais clairement différentes de celles des régressions PLS et RSPP, à en juger par la faiblesse des corrélations (de l'ordre de 0.5). Le décideur doit donc être conscient qu'il obtiendra des poids très différents s'il opte pour la RC ou la méthode univariée de préférence à la PLS ou la RSPP.

Si l'on exclut de la régression en cascade les trois dimensions de satisfaction en les remplaçant pas des

variables latentes, tout en gardant le même arbre, les poids se rapprochent de ceux de la régression séquentielle pas-à-pas ( $r=0.91$ ) et, dans une moindre mesure, ceux de la PLS ( $r=0.71$ ). A périmètre de variables identique, la RSPP produit, sans définition préalable de l'arbre, des poids comparables à ceux de la régression en cascade. Le recours à cette dernière ne se justifie que si l'on cherche à reconstituer la satisfaction globale sur les dimensions intermédiaires plutôt que sur l'ensemble des items.

#### *Stabilité des poids*

Pour le décideur, la stabilité des poids est plus importante que la précision de l'ajustement. Même si elle n'affecte qu'une minorité de poids, la dispersion des poids d'un échantillon à l'autre est perturbante.

Globalement, les résultats sont assez stables : la précision n'est que légèrement dégradée en prédictif sur l'échantillon test. Mais cette stabilité globale masque des écarts fréquents et parfois importants : des écarts de plus de 30% entre les poids des deux échantillons s'observent pour 1 cas sur 2 en PLS et cascade, et 1 cas sur 3 en RSPP. En deux occasions (une seule pour la cascade), les poids varient même du simple au double.

< insérer ici le tableau 3 >

Des trois méthodes multivariées, la régression séquentielle pas-à-pas offre ici le meilleur compromis entre les deux points de vue sur la stabilité. Mais c'est la méthode univariée qui est, de loin, la plus stable : les poids de travail et de test sont corrélés à hauteur de 0.97 et aucun des 25 poids ne diffère d'un échantillon à l'autre de plus de 30% en valeur absolue. Cette stabilité explique la préférence de certains décideurs pour ce type de méthode.

#### *Dispersion des poids*

Le principal inconvénient de la méthode univariée est de donner trop de poids à l'importance partagée par rapport à l'importance spécifique. D'où une faible dispersion des poids, à la fois globalement (écart-type faible) et aux extrêmes (rapport de 1 à 3 entre les poids minima et maxima, contre 1 à 10 pour les méthodes multivariées). Si l'on en juge par le niveau des corrélations entre les méthodes (de 0.2 à 0.5), il est clair qu'on ne mesure pas la même chose en univarié et en multivarié.

En régression PLS, 5 poids sur 29 sont négatifs ce qui conduit à la plus grande dispersion observée (le plus fort écart-type des poids). En valeur absolue, cette différence de dispersion disparaît. Hormis cette inversion indésirable de signe, ces poids sont d'une importance comparable à celle de la régression séquentielle. La régression en cascade tend également à doter une minorité de critères de poids très faibles. Là encore, la régression séquentielle pas-à-pas fait preuve du meilleur équilibre.

#### *Implications managériales*

Tire-t-on des conclusions différentes des hiérarchies de

poids selon que l'on utilise une méthode de calcul plutôt qu'une autre ? L'examen des poids reproduits dans les tableaux 4 et 5 apporte quelques éléments de réponse à une question souvent posée par les décideurs (pour une question de confidentialité, les libellés des items ne sont pas révélés).

< insérer ici le tableau 4 >

Un consensus s'établit pour les critères de satisfaction à importance très faible (items 21 à 25). Pour ce type de critère, le choix de la méthode importe peu.

Les critères à importance forte sont identifiés comme tels par les régressions PLS et RSPP (items 1 à 6). En revanche, trois critères jugés importants par les régressions PLS et RSPP (items 1, 3 et 7), sont dotés d'un poids faible par la régression en cascade, et vice-versa, certains items jugés importants par la régression en cascade (ex. : items 15, 16 et 18) le sont moins en régression PLS ou RSPP. Il est clair que l'on tirera des conclusions différentes de ces données, selon que l'on utilisera les régressions PLS ou RSPP d'une part, ou la régression en cascade, d'autre part.

Pourquoi une telle différence entre les poids des items, d'un type de méthode à l'autre ? Elle semble, ici, provenir d'une divergence d'appréciation du poids respectif des dimensions par la régression en cascade.

< insérer ici le tableau 5 >

Le poids de la dimension « satisfaction à la livraison » est beaucoup plus faible en RC qu'en PLS ou RSPP. Rappelons que ce poids est obtenu directement en RC (coefficient de régression au noeud final), et indirectement en PLS/RSPP (simple addition des poids des items constitutifs de la dimension). En RC, l'écart de poids entre les dimensions se répercute sur les items en proportion de leur lien avec la dimension. Si l'item représentant la dimension est faiblement explicatif de la satisfaction globale, alors, mécaniquement, les items constitutifs de cette dimension seront eux-aussi faiblement explicatifs de la satisfaction globale, indépendamment de leur propre corrélation avec cette satisfaction. Il faut donc veiller à ce que l'item représentant la dimension soit à la fois plus explicatif de la satisfaction globale que chacun des items constituant cette dimension, mais aussi qu'il représente ces items au mieux. Cela ne semble pas être suffisamment le cas ici, où l'on constate par comparaison de méthodes que les items constitutifs de la dimension « satisfaction à la livraison » sont plus explicatifs de la satisfaction globale que l'item représentatif de cette dimension. D'où l'intérêt d'une éventuelle reformulation des questions pour équilibrer au mieux le poids des dimensions, si l'on se fie aux résultats des régressions PLS/RSPP.

La différence entre les régressions PLS et RSPP porte surtout sur les critères d'importance moyenne. Globalement, la régression PLS a privilégié les critères liés à la voiture (ex. : *conformité du véhicule*, *bon fonctionnement du véhicule*), la régression RSPP, elle, privilégiant les critères relationnels (ex. : *amabilité et*

*disponibilité, capacité à fournir des informations précises*). Toutefois, si les procédures mises en œuvre par ces deux méthodes pour minimiser les conséquences indésirables de la multicollinéarité, conduisent in fine à une lecture parfois différente des données (cf. items 7, 8 et 9), force est de conclure qu'elles produisent approximativement les mêmes hiérarchies.

Finalement, la présence de poids négatifs en régression PLS, est explicable. Il s'agit de critères périphériques à la satisfaction (ex. : *aspect des locaux, variété des véhicules exposés*), très liés à la taille de la concession. Certaines petites concessions ne paient pas de mine, mais offrent à leurs clientèle une qualité relationnelle de premier ordre. A l'inverse, des concessions grandes, modernes et très bien équipées, peuvent déconcerter une clientèle plus attachée à la qualité de la relation qu'à l'aspect des locaux. D'où une relation inverse entre la satisfaction globale et l'aspect des locaux, par exemple, à qualité des travaux constante. Ces phénomènes modérateurs sont malheureusement difficiles à expliquer et à justifier dans le cadre de baromètres de satisfaction largement diffusés au sein de l'entreprise.

### **Conclusion et Voies de recherche**

L'étude empirique, représentative des résultats habituellement obtenus, conduit à plusieurs conclusions.

Aucune méthode n'offre une supériorité nette pour la qualité de l'ajustement, celle-ci restant assez peu satisfaisante, même s'il s'agit de données individuelles. L'incapacité relative à reconstituer l'appréciation globale à partir d'un inventaire a priori exhaustif des facteurs censés l'influencer, résulte de la conjonction de plusieurs facteurs : une clientèle hétérogène (*ce qui est important pour l'un ne l'est pas pour l'autre*), un questionnaire trop détaillé au goût des clients (*« pourquoi toutes ces questions ! »*) et une absence d'opinion tranchée (*d'où le caractère aléatoire de certaines réponses*). Ce constat appelle un effort de recherche accru pour améliorer la démarche à chaque niveau : questionnaire, échelles, mode de recueil et traitement de l'information.

Les résultats des méthodes univariées et multivariées sont clairement différents. Si l'on recherche la meilleure stabilité des poids dans le temps ou d'un échantillon à l'autre, la méthode univariée est clairement indiquée. Cette stabilité, toutefois, est obtenue au détriment de la dispersion des poids, peu différenciés. Elle conduit souvent à considérer que tout, ou presque, est important, ce qui ne facilite pas l'identification des priorités. Pour une interprétation des poids, la méthode univariée n'est donc pas recommandée.

La régression en cascade fournit des résultats différents de ceux des régressions PLS et RSPP. Elle est donc plus complémentaire que concurrente de ces dernières. Il s'agit d'une méthode d'estimation des paramètres d'un modèle hiérarchique qui se veut explicatif de la formation de la satisfaction globale ; au contraire, les deux autres méthodes ne font pas d'hypothèses a priori

et sont de nature exploratoires et descriptives. Si les données sont de nature hiérarchique, et si par surcroît le nombre de critères est élevé, la régression en cascade est certainement plus adaptée que la régression PLS. Son aptitude à prendre en compte toute l'information disponible (dimensions, thèmes, sous-thèmes etc.) dans un modèle structuré, à l'inverse de l'approche purement empirique des régressions PLS et RSPP, et le fait qu'elle accorde plus d'importance aux critères intermédiaires qu'aux critères de base, sont à la fois un atout et une source de difficulté.

En effet, dès lors qu'un critère de satisfaction s'intègre mal à la dimension à laquelle il est affecté, son poids est minoré. Inversement, à pouvoir explicatif de la satisfaction globale constant, une forte corrélation avec le critère de niveau supérieur favorise le poids. D'où l'importance de la bonne construction de l'arbre et de la constitution de nœuds de critères les plus homogènes possibles. La sensibilité des résultats à l'arbre choisi reste cependant une difficulté sérieuse de mise en œuvre de cette méthode, tant que n'aura pas été définie une démarche moins empirique de construction et de test d'arbres hiérarchiques.

Un dernier constat est la convergence des résultats de la régression PLS et de la régression séquentielle pas-à-pas, ainsi que la dominance empirique de cette dernière. Elle fournit des résultats proches de ceux de la régression en cascade lorsque les variables intermédiaires sont latentes et tient la comparaison avec les résultats de la régression PLS. La confrontation des poids obtenus est intéressante lorsque l'on prend en compte la séquence dans laquelle chaque critère est entré (graphique 5).

< Insérer ici le graphique 5 >

Graphiquement, les méthodes convergent de manière satisfaisante à deux exceptions près : d'une part, les critères mineurs de la première séquence ont un poids relatif plus faible qu'en régression PLS ; d'autre part, les critères de la première séquence ont un poids trop important par rapport à ceux des autres séquences. Corrigé de ces deux « anomalies » par une variable binaire, la corrélation entre les deux séries de poids atteint 0.96. Dans l'attente d'une régression PLS sous contrainte, nous encourageons donc les chercheurs à se pencher sur le problème de la détermination a priori, et non a posteriori comme ici, du coefficient correcteur qui permettrait de faire rapprocher les poids de la régression séquentielle de ceux de PLS. Dans la mesure où la régression séquentielle garantit le respect du signe des poids et gère beaucoup mieux la redondance que la régression PLS, cette amélioration aurait un intérêt pratique indéniable. Elle offrirait au praticien satisfait de la régression PLS, une alternative acceptable lorsque le nombre de variables devient trop important pour elle. Plus les résultats de la régression RSPP convergeront vers ceux de la PLS lorsque la comparaison est possible, plus grande sera la confiance du décideur dans les résultats de la RSPP lorsqu'elle sera la seule méthode applicable. Cela, parce que la régression PLS jouit d'une



**Tableau 1** : Performances des méthodes

<b>Régression PLS (2 composantes)</b>				
Echantillon	Travail		Test	
Coefficients	Travail	Test	Travail	Test
R <sup>2</sup>	0,496	0,480	0,469	0,483
RMSE	9,03%	9,17%	9,62%	9,49%
<b>Régression séquentielle pas à pas</b>				
Echantillon	Calibration		Test	
Coefficients	Travail	Test	Travail	Test
R <sup>2</sup>	0,494	0,467	0,466	0,479
RMSE	9,05%	9,30%	9,74%	9,52%
<b>Régression en cascade</b>				
Echantillon	Travail		Test	
Coefficients	Travail	Test	Travail	Test
R <sup>2</sup>	0,536	0,524	0,590	0,595
RMSE	8,82%	8,84%	8,51%	8,48%

R<sup>2</sup> : pourcentage de variance expliquée

RMSE : erreur quadratique moyenne rapportée à la moyenne

**Tableau 2** : corrélation entre les poids des critères par méthode et échantillon

PLS (travail)	PLS (test)	RSPP (travail)	RSPP (test)	Cascade (travail)	Cascade (test)	Univarié (travail)	Univarié (test)
PLS (travail)	<b>0,85</b>	0,84	0,71	0,52	0,44	0,55	0,50
	PLS (test)	0,75	0,82	0,54	0,54	0,37	0,43
		RSPP (travail)	<b>0,80</b>	0,64	0,58	0,38	0,35
			RSPP (test)	0,60	0,62	0,42	0,46
				Cascade (travail)	<b>0,88</b>	0,17	0,21
					Cascade (test)	0,24	0,33
						Univarié (travail)	<b>0,97</b>

**Tableau 3** : Stabilité et variabilité des poids

	<b>Stabilité (sur 25)</b>	<b>Variabilité</b>	
		Ecart- type	Nb poids < 10% du maxi
PLS (travail/test)	56%	40,5	5
RSPP (travail/test)	32%	27,3	1
Cascade (travail/test)	48%	24,7	4
Univarié (travail/test)	0%	19,6	0

Mesure de stabilité : proportion de poids de test supérieurs de 30% en valeur absolue aux poids de travail.

Mesure de variabilité des poids (travail) : écart-type des poids

**Tableau 4 :** Poids des critères de satisfaction

	PLS	RSPP	Cascade	Univarié
Item n°1	7,73	6,56	3,63	4,70
Item n°2	7,38	8,90	12,90	2,54
Item n°3	7,17	7,13	2,55	4,17
Item n°4	6,68	7,61	10,14	5,15
Item n°5	6,62	6,26	7,99	5,07
Item n°6	6,28	8,20	6,71	5,38
Item n°7	5,71	3,32	1,43	4,94
Item n°8	5,41	2,56	6,11	3,30
Item n°9	4,97	3,08	1,90	5,46
Item n°10	4,87	4,22	2,16	4,54
Item n°11	4,73	6,61	7,36	3,81
Item n°12	4,46	5,96	0,75	4,19
Item n°13	4,29	3,17	2,39	4,36
Item n°14	3,49	3,92	3,80	2,41
Item n°15	2,68	3,27	4,25	5,28
Item n°16	2,58	3,19	5,33	5,53
Item n°17	2,50	2,53	0,39	3,99
Item n°18	1,86	2,76	6,73	4,71
Item n°19	1,61	1,54	3,92	3,68
Item n°20	1,10	3,36	1,60	3,69
Item n°21	-0,68	0,81	2,00	2,15
Item n°22	-1,43	1,74	1,10	3,04
Item n°23	-1,60	1,57	0,85	2,24
Item n°24	-2,00	0,90	1,61	3,02
Item n°25	-2,17	0,85	2,41	2,63
Total (valeur absolue)	100,00	100,00	100,00	100,00

**Tableau 5 :** Poids des dimensions de satisfaction

	PLS	RSPP	Cascade
Commande	40	41	58
Livraison	53	51	34
Locaux	7	8	8
Total	100	100	100

### Notes

<sup>i</sup> Cette stratégie s'apparente à la pratique des échelles psychométriques consistant à sommer le score des items censés mesurer le même concept. Lorsque les items d'un même nœud mesurent la même chose,  $r_{ij} \rightarrow 1$  et  $r_i \approx r_j = r$ , d'où un poids égal à  $r/n$ .

<sup>ii</sup> Le poids des variables d'un modèle hiérarchique dépend de la structure de l'arbre, c-à-d du nombre de niveaux hiérarchiques, du nombre de nœuds et du nombre de variables à l'intérieur de ces nœuds. Windal [6] a proposé un coefficient correcteur facilitant la comparaison entre variables appartenant à des nœuds de tailles très différentes. Si  $P_{ik}$  dénote le poids non corrigé,  $W_{ik}$  le poids corrigé,  $P_i$  la somme des poids non corrigés,  $N$  le nombre d'items du nœud  $k$ , alors  $W_{ik} = N \cdot [1 - (\sigma(P_{ik}) \cdot \sqrt{(N-1)})/P_i] \cdot P_{ik}$ .

<sup>iii</sup> Si  $S$  dénote la satisfaction globale, chaque nœud privé de variable dépendante donne lieu à deux opérations successives : (1) Calcul du poids des items dans le nœud (sur variables centrées-réduites) :  $S = a_1(A) + a_2(NC) + \varepsilon$  et (2) Remplacement de «  $X$  » par  $\hat{S}$ , la satisfaction reconstituée :  $X = \hat{S} = a_1(A) + a_2(NC)$

<sup>iv</sup> Le coefficient de régression standardisé est égal au produit du coefficient non standardisé par le ratio des écarts-types du prédicteur et de la variable à expliquer.

<sup>v</sup> Si une variable est dotée d'un coefficient de régression statistiquement nul, c'est que son coefficient de corrélation simple l'est également, puisqu'elle est seule dans « sa » régression. En revanche, une variable peut avoir un poids non "nul" même si la corrélation simple est "nulle" au départ.

<sup>vi</sup> Les regroupements des items suggérés par l'ACP conduisent en général à des arbres dotés d'un nombre important de coefficients de régression non significatifs. Trop d'items participent à la première composante, d'où un déséquilibre dans le pouvoir explicatif des nœuds. En régression en cascade, on cherche au contraire à construire des nœuds équilibrés pour minorer les effets de la colinéarité en amont de l'arbre.

<sup>vii</sup> Les hiérarchies des poids avec et sans les trois items correspondant aux dimensions de satisfaction, sont très différentes. Ce sont les items les plus corrélés aux trois dimensions, donc dotés de poids élevés en leur absence, qui souffriraient le plus de la prise en compte inopportune des dimensions.

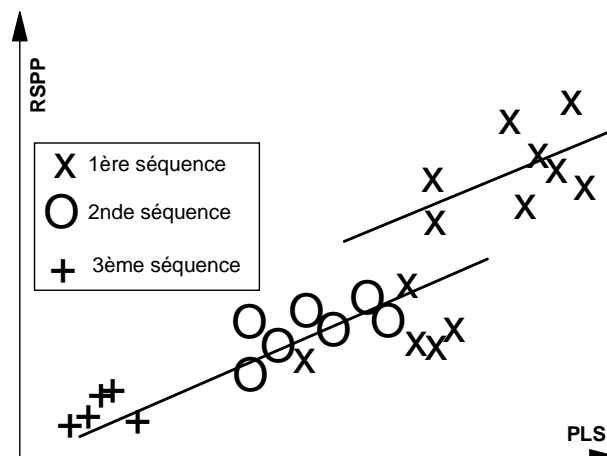


Figure 5 : Comparaison des poids PLS et RSPP