

PREVISION DES VENTES

Polycopié du produit multimédia

**Auteur : Régis Bourbonnais
Université de Paris-Dauphine**

Octobre 2001

Sommaire

Chapitre 1 : LA PREVISION DES VENTES : QUELQUES REFLEXIONS PRELIMINAIRES

1. Pourquoi élaborer des prévisions des ventes ?.....	4
1.1. La prévision conditionne l'optimisation.....	4
1.2. L'obligation de prévoir est liée aux délais de réaction.....	4
2. Ce qui caractérise un problème de prévision.....	5
3. Qui prévoit ?.....	6
3.1. Le processus d'élaboration de la prévision.....	6
3.2. La prévision comme une aide à la décision.....	7

Chapitre 2 : LA DECOMPOSITION D'UN HISTORIQUE DES VENTES

1. La notion d'historique	9
2. Le calcul des statistiques de base et la présentation des données	10
2.1. Le calcul des statistiques de base.....	10
2.2. La présentation des données	11
3. L'analyse de série chronologique et la prévision	14
3.1. Les étapes de la prévision des ventes.....	14
3.2. Recherche et correction de valeurs anormales.....	15
3.2.1 La détection des valeurs anormales	15
3.2.2 La correction des valeurs anormales.....	17
3.3. Le schéma de décomposition	18
3.4. La tendance	20
3.5. La saisonnalité.....	22
3.5.1 Pourquoi analyser une saisonnalité ?	22
3.5.2 La correction des jours ouvrables	22
3.5.3 Le calcul de la saisonnalité	23
3.5.4 Test sur la saisonnalité	30
3.5.5 Famille de saisonnalité.....	32
3.6. La composante cyclique	33
3.7. Le facteur résiduel.....	33

Chapitre 3 : PRINCIPALES METHODES (1) : LES METHODES FONDEES SUR LE LISSAGE EXPONENTIEL

1. Caractéristiques des méthodes de lissage.....	34
1.1. Principes de base.....	34
1.2. Formulation.....	34
1.3. Rôle de la constante de lissage.....	36
2. Le lissage simple : le modèle stationnaire	38
3. Le lissage exponentiel double : le modèle linéaire	39
4. Les modèles de Holt et Holt–Winters.....	42
4.1. Le modèle de Holt.....	42
4.2. Le modèle avec tendance et saisonnalité (modèle de Holt–Winters)	42

5.	Comment choisir le coefficient de lissage ?.....	45
5.1.	Principes généraux	45
5.2.	Valeur de α qui minimise la somme des carrés des erreurs de prévision.....	45
5.3.	Procédure de régulation d'un coefficient de lissage.....	46
5.3.1	Principes généraux	46
5.3.2	Les variables de contrôle.....	47
5.3.3	Application.....	50

Chapitre 4 : PRINCIPALES METHODES (2) : BOX-JENKINS ET MODELE CAUSAL

1.	La corrélation et le corrélogramme	54
1.1.	La corrélation	54
1.1.1	Mesure du coefficient de corrélation linéaire	54
1.1.2	Limites de la notion de corrélation	56
1.2.	Le corrélogramme et la fonction d'autocorrélation.....	56
1.3.	Introduction à la méthodologie de Box et Jenkins	60
1.3.1	Typologie des modèles.....	60
1.3.2	Le problème de la stationnarité.....	62
1.3.3	Recherche de la représentation adéquate	62
2.	L'approche exogène	65
2.1.	La notion d'ajustement et le modèle de régression simple.....	66
2.2.	La modélisation à plusieurs variables : l'économétrie	67
2.3.	Interprétation statistique d'un modèle	68
2.4.	Exemples d'estimation économétrique à l'aide des outils sur tableur Excel.....	71
2.5.	La sélection de variables explicatives.....	74
1.	Mesure de la qualité d'une prévision.....	79

Chapitre 5 : EVALUATION DES METHODES DE PREVISION

1.1.	Les indicateurs de mesure.....	79
1.2.	La prévision est-elle biaisée ?	80
1.3.	Convergence de la prévision.....	81
1.4.	Détection de la cause de l'erreur	82
1.5.	Quelle erreur de prévision est admissible ?	83
1.6.	La simulation.....	83
2.	Evaluation synthétique de la qualité de la prévision.....	84
2.1.	Calcul d'un indice pondéré.....	84
2.2.	Evaluation graphique	85
3.	Procédures d'évaluation	87
3.1.	Durée de la mesure.....	87
3.2.	Comparaisons de prévisions	88
3.3.	Exemples d'outils d'alerte	89

Chapitre 1

LA PREVISION DES VENTES : QUELQUES REFLEXIONS PRELIMINAIRES

Nous examinons dans ce chapitre introductif en quoi l'obligation de prévoir la demande dépend de l'organisation même de l'entreprise et de son secteur d'activité. Nous présentons ensuite comment aborder la prévision des ventes dans l'entreprise : comment la caractériser en terme de problème et de solution. Puis nous proposerons quelques définitions permettant de distinguer les notions de prévision et d'objectif. Enfin nous montrons que ce problème de prévision est transversal par rapport aux fonctions de l'entreprise.

1. Pourquoi élaborer des prévisions des ventes ?

Cette question peut sembler triviale, cependant elle mérite réflexion et s'avère plus complexe qu'il n'y paraît dans sa réponse. Deux points sont abordés.

1.1. La prévision conditionne l'optimisation

La prévision de la demande est à la base de l'optimisation de la chaîne logistique. Quel que soit le type d'organisation d'une entreprise (flexibilité et réactivité) et son degré d'intégration dans le pilotage des flux, deux éléments sont essentiels pour déterminer le calcul d'un approvisionnement optimal :

- la prévision de consommation,
- la fiabilité attendue de cette prévision.

La prévision de consommation détermine de manière directe une partie du niveau d'approvisionnement ; il s'agit de couvrir au minimum le besoin pour un certain délai.

La fiabilité attendue de cette prévision, ou la crédibilité de la prévision, permet de dimensionner de manière optimale le niveau du stock de sécurité.

1.2. L'obligation de prévoir est liée aux délais de réaction

Toutes les entreprises ne sont pas dans l'obligation de prévoir, du moins dans le cadre de leur système de pilotage des flux.

Le cas extrême est représenté par les entreprises qui travaillent « à la commande », c'est-à-dire celles dont le délai de livraison accepté par ses clients est supérieur au délai d'approvisionnement des matières auprès des fournisseurs et au délai de production. Quelques entreprises sont dans ce cas : notamment celles du secteur aéronautique, des travaux publics, des chantiers navals. Pour elles, le problème de la prévision des ventes, à court et moyen terme, ne se pose pas.

A l'opposé, figure la grande majorité des entreprises du secteur de la grande consommation. Le délai de livraison est de quelques jours, voire de quelques heures (répartiteur pharmaceutique) ; le stockage des produits finis est alors impératif sous peine d'être en rupture. Les conséquences sont le risque de la perte de la vente et la dégradation de l'image de marque.

Pour optimiser le niveau de stock, l'obligation de prévoir la demande s'impose.

Le Tableau 1 illustre l'arbitrage que l'entreprise est en mesure d'effectuer : flexibilité ou stockage. Souvent, on cherche à résoudre un problème de prévision qui peut trouver sa solution plus facilement par un accroissement de la réactivité et/ou de la flexibilité. Cependant, ce choix souvent n'existe pas pour l'entreprise, il est imposé par les délais de livraison que les fournisseurs exigent.

Tableau 1 – Pourquoi prévoir ?

Approvisionnement matière	Fabrication	Emballage personnalisation		Stock
à la commande	à la commande	à la commande	⇒	Aucun
<i>sur prévision</i>	à la commande	à la commande	⇒	matière
<i>sur prévision</i>	<i>sur prévision</i>	à la commande	⇒	semi-finis
<i>sur prévision</i>	<i>sur prévision</i>	<i>sur prévision</i>	⇒	produits finis

Quelques définitions préliminaires

2. Ce qui caractérise un problème de prévision

La première approche de la prévision consiste à en mesurer les spécificités. Pour envisager une typologie des problèmes de prévision en terme d'horizon, de type de produit ou de secteur, ou en terme de but opérationnel, il est utile de dégager quelques traits qui différencient fondamentalement :

- le secteur d'activité,
- l'utilisation opérationnelle,
- la (ou les) fonction(s) utilisatrice(s) de la prévision,
- l'horizon.

L'approche est très dépendante du secteur d'activité : on ne prévoit pas des livraisons de ciment par les mêmes méthodes que des ventes de savons. Les causalités économiques sous-jacentes sont différentes suivant que le secteur est plus ou moins en amont dans le circuit industriel donc plus ou moins proche de la demande finale, suivant que le produit est stockable ou non, qu'il donne lieu à un marché de renouvellement (télévision) ou non (acier), que le produit est standardisé ou non. Tous ces éléments jouent un rôle dans la définition d'une typologie de trois grands secteurs, qui dans cet ouvrage sont étudiés dans les différents chapitres de la deuxième partie (approche sectorielle).

Nous pouvons distinguer :

- le domaine des biens industriels intermédiaires (acier, produits chimiques, matériaux de construction) qui dépendent étroitement de l'environnement économique général, et de l'évolution de la demande dans de grandes branches (bâtiment, automobile ...).

L'influence de la firme, par sa politique volontariste sur le marché, est assez limitée : le problème du prévisionniste est alors de savoir quand se passera le retournement de conjoncture ;

- le domaine des produits de grande consommation est à l'extrême opposé : l'influence de la conjoncture générale est relativement faible ; ceci est largement compensé par une forte sensibilité à l'environnement concurrentiel. Le prévisionniste doit déterminer quel sera le volume des ventes compte tenu de l'action marketing de la firme et de celle de ses concurrents ;

- à mi-chemin se situent les biens de consommation durables (automobile, électro-ménager, télévision, etc.). Sensibles à la fois à la conjoncture économique générale et à des variables marketing (efforts promotionnels, effets de mode, par exemple), ils sont l'objet de méthodes de prévision différentes. Ce qui caractérise ces marchés est la présence juxtaposée d'une demande de premier équipement et d'une demande de renouvellement.

3. Qui prévoit ?

3.1. Le processus d'élaboration de la prévision

La responsabilité d'élaboration de la prévision est très variable d'une entreprise à l'autre.

Quelques phrases issues d'interviews en entreprise sur ce sujet :

« Les commerciaux sont les mieux placés car ils sont en contact permanent avec les clients, cependant leur manque de recul et, surtout, de motivation par rapport aux travaux administratifs font que nous ne tenons plus compte de leur avis ».

« Les hommes de marketing connaissent bien leur marché, malheureusement ils confondent leurs désirs avec la réalité ».

« Les logisticiens sont isolés du marché et des clients. Ils ont l'œil rivé sur les stocks et préfèrent la rupture au sur-stock. Cependant, par leur habitude de mesurer les flux de sortie, ils ont un sens aigu des chiffres. Ce sont les seuls réellement obligés à élaborer des prévisions ».

Ces trois affirmations sont révélatrices de l'état d'esprit de chacun.

La solution traditionnelle est de laisser à l'homme du terrain (vendeur, représentant, directeur régional, etc.) le soin de réaliser des prévisions. Les arguments favorables sont nombreux : en contact perpétuel avec le produit et les acheteurs, il a le sens du marché ; il est également le premier intéressé à bien prévoir puisqu'il sera jugé sur l'écart entre sa réalisation et sa prévision.

Mais les défauts sont manifestes :

– d'une part cela revient, comme nous venons de le voir au point précédent, à confondre prévision et objectif, l'homme de terrain faisant un mélange entre ce qu'il suppose de l'évolution du marché et ce qu'il pense pouvoir faire auprès de la clientèle ;

– d'autre part, il existe une forte tendance à la prévision « marginale ». Il suffit que le dernier client, rencontré avant d'émettre la prévision, ait passé une commande importante, ou que le jour précédent ait été très faste, pour que la prévision pour les 3 ou 6 mois à venir devienne très optimiste. Et inversement.

L'expérience prouve que la prévision des hommes de terrain est souvent biaisée (les écarts sont systématiquement positifs ou négatifs), pour deux raisons :

– le tempérament personnel du prévisionniste intervient non seulement suivant qu'il est optimiste ou pessimiste, mais aussi suivant sa préférence pour des écarts potentiels positifs ou négatifs. Ainsi un vendeur peut, par prudence, émettre des prévisions très « conservatrices », alors qu'un autre ne craindra pas les écarts négatifs, souhaitant au contraire être « entraîné » par sa prévision ;

– le système de gestion prévisionnelle biaise très sensiblement les prévisions des individus. Si par exemple les moyens sont affectés aux directions régionales de vente en fonction de la prévision, ils auront tendance à l'optimisme. Si les primes sont liées à la réalisation de l'objectif, ils auront tendance au pessimisme.

Nous appellerons prévision entre guillemets (« prévision ») ce type de données bâtarde, à mi-chemin entre un objectif réel et une véritable prévision. Leur mauvaise qualité ne met pas en cause la performance des hommes du terrain et leur capacité à effectuer un apport essentiel dans un système de prévision : elle condamne tout système de gestion prévisionnelle qui biaise d'emblée les données d'entrée qui lui sont nécessaires.

Souvent lorsque des écarts trop importants apparaissent sur les « prévisions », l'entreprise est tentée de ne plus accorder sa confiance aux hommes de terrain, qui se voient ainsi disqualifiés par rapport à une tâche qui ne leur semblait pourtant pas la moins intéressante. En conséquence, l'entreprise recourt à des profils entièrement opposés : la prévision devient l'affaire de statisticiens, qui ne seront pas suspects d'implication dans le fonctionnement interne de la firme. On assiste alors

au développement de ce que l'on peut nommer familièrement « la méthode des matheux » : l'application aveugle d'une méthodologie statistique à des prévisions de ventes en est une bonne illustration. Les ventes deviennent une série chronologique, suite abstraite de nombres, à laquelle il importe d'ajuster le meilleur modèle.

L'éloignement de toute référence concrète aboutit à deux conséquences :

– une autosatisfaction méthodologique : le souci de la modélisation (comment identifier la structure du modèle ? quelle méthode d'estimation statistique utiliser ? etc.) va rapidement déborder et provisoirement dominer celui de la prévision. Le but, bien prévoir, disparaît devant les moyens de l'atteindre, la qualité du modèle statistique est une fin en soi. Si la liaison entre les deux était biunivoque et directe, cela ne poserait pas de problème. Un bon modèle au sens statistique donnerait automatiquement une bonne prévision. Malheureusement l'expérience prouve que parfois tel n'est pas le cas ;

– une coupure avec les utilisateurs, due à l'absence de langage commun, et à l'éloignement des rationalités, mathématique et explicite pour les uns, intuitive et implicite pour les autres. Cette coupure provoque un scepticisme face aux modèles dès que le moindre écart se révèle entre prévisions et réalisations.

L'utilisation d'une pure méthodologie statistique a rarement pu aboutir à l'installation durable d'un système de prévision dans une entreprise. Entre ces deux écueils, que faire ? Même si cela peut apparaître difficile aux uns et aux autres : développer une coopération entre ces deux approches, en établissant des systèmes de prévision qui ne sont pas simplement des modèles de traitement statistique, mais qui associent les utilisateurs en intégrant des variables significatives par rapport à leurs schémas de référence.

Le plus souvent, c'est la fonction logistique qui est en charge de la prévision ; et pourtant les hommes de logistique ne sont pas les mieux placés du fait de leur isolement vis à vis du marché. Mais l'obligation de fournir des prévisions quantitatives et fines aux fournisseurs ou à la production fait qu'ils sont jugés responsables en cas de problème.

De manière idéale, l'élaboration de la prévision doit se faire en commun au sein d'un comité réunissant l'ensemble des fonctions citées précédemment. Chacun doit s'engager sur des chiffres et à ce titre en est responsable. Ainsi, le risque de focaliser tous les problèmes – ruptures ou surstocks – auprès d'une unique personne, « le prévisionniste », est très fortement diminué.

L'objectif est de disposer, au sein d'un même outil, de toutes les informations susceptibles d'améliorer le processus d'élaboration de la prévision des ventes.

Des capteurs « informations terrains » (vendeurs) indiquent, via un micro-ordinateur portable, leur prévision en fonction des actions promotionnelles planifiées ainsi que toutes informations susceptibles d'influencer les ventes (nouveau client, référencement, promotion client ...). Ces informations sont consolidées au niveau central et directement accessibles par le responsable commercial. Cette prévision est ensuite confrontée à la prévision marketing et à la prévision statistique lors du « Comité Prévision ».

3.2. La prévision comme une aide à la décision

La vision que l'on a de l'apport du prévisionniste est souvent un peu magique : elle s'appuie sur un besoin irrationnel qui prend ses racines dans des époques reculées. La divination était à l'honneur dans les civilisations méditerranéennes avant l'ère chrétienne : le christianisme mit fin aux méthodes divinatoires qu'il considérait comme liées aux cultes païens. Ainsi, Saint Augustin se dressa contre l'illusion des horoscopes et arriva à la conclusion suivante : « Il n'y a donc point d'art divinatoire. C'est par hasard qu'on prédit vrai ».

L'utilisation directement opérationnelle de la prévision n'est pas souhaitable : elle apparaît plus comme un outil d'aide à la décision qu'un substitut complet de la réflexion personnelle. L'utilisation

directe, sans une étape de validation, de données prévisionnelles serait dangereuse même si parfois cette étape peut être réduite comme par exemple pour une application de gestion de stock et d'approvisionnement. De plus c'est la conjonction de l'utilisation d'un système de prévision formalisé et de la capacité de réflexion du prévisionniste qui permet d'améliorer notablement la qualité finale de la prévision.

La prévision a doublement hérité des techniques divinatoires : d'une part elle satisfait le même besoin irrationnel et immense de connaître l'avenir, d'autre part elle se pare d'un côté magique en utilisant l'outil mathématique, ésotérique comme l'était l'hépatoscopie ou l'astrologie, même si elle justifie son efficacité par la rationalité de l'outil. L'utilisateur attend beaucoup de la prévision ; cet espoir constitue pour le prévisionniste à la fois un atout et un risque : l'intérêt qui existe lors de la mise en place du système peut être suivi d'une déception si les premiers résultats ne correspondent pas à l'attente.

L'apport d'un système de prévision se situe à deux niveaux :

– dans la durée – c'est-à-dire à la suite de nombreuses émissions de prévisions – pour sa capacité à réduire significativement la marge d'incertitude. Exprimé en termes statistiques, on pourra considérer comme une bonne performance de réduire de moitié l'écart-type de l'erreur de prévision ;

– ponctuellement, savoir anticiper les retournements de conjoncture. Ou alternativement : savoir prédire si un retournement observé récemment a un caractère durable ou non.

L'entreprise peut attendre beaucoup de la prévision à condition de bien l'utiliser. Il faut à tout prix éviter la vision passéiste qu'entraîne le sentiment que la prévision est ce que la firme subira inévitablement. Elle est bien au contraire un outil d'aide à la décision : elle doit servir à l'entreprise à agir sur son futur et non à le subir.

Chapitre 2

LA DECOMPOSITION D'UN HISTORIQUE DES VENTES

Dans ce chapitre, nous allons aborder ce qui caractérise d'un point de vue statistique les données servant de base à l'étude prévisionnelle, en examinant la notion d'historique et de série chronologique. L'examen des données est le préalable à toute étude quantitative, que ce soit dans un but d'analyse ou de prévision.

Nous traiterons ensuite en détail de l'analyse de série chronologique : Qu'est-ce qu'une observation anormale ? Comment la corriger ? Comment calculer une tendance ? Comment retirer une saisonnalité ? En définissant un schéma de décomposition, nous pouvons retirer d'un historique les effets répétitifs qui peuvent être facilement extrapolés, afin de dégager un terme résiduel représentant les fluctuations erratiques de la chronique.

1. La notion d'historique

Une série chronologique est composée d'un historique, c'est-à-dire d'une suite de valeurs ordonnées dans le temps à périodicité constante : par exemple, l'indice mensuel des prix à la consommation de la France entière publié par l'INSEE en base 100 en 1980. La « suite » des valeurs de janvier 1995 jusqu'à septembre 2000 au mois le mois représente l'historique de l'indice des prix à la consommation. Les ventes mensuelles, pour un article, de mars 1997 à février 2000 serviront de base historique afin d'élaborer la prévision.

Un historique doit réunir certaines propriétés :

- *Il ne se compose que de valeurs connues et calculées*, qui sont effectivement réalisées.
- *Il est représentatif de ce que l'on cherche à prévoir*. Un historique des livraisons ne permet pas de prévoir la demande !
- *Il est homogène dans le temps*. Pour reprendre notre exemple d'indice de prix, toutes les valeurs sont en base 100 en 1980 et correspondent au même mode de calcul et de collecte statistique.
- *Il comprend un nombre minimal d'observations*. Il semble logique d'écrire que plus l'historique est long, meilleure sera la qualité de l'analyse et par voie de conséquence de la prévision. La longueur minimum d'un historique, c'est-à-dire son nombre de valeurs, est variable selon le but recherché et la périodicité des données.
Si l'objectif est de fournir des prévisions de périodicité mensuelle à un horizon de 6/12 mois, nous considérons alors que le minimum de valeurs disponibles est de 3 ans, soit 36 observations mensuelles, alors que la dimension souhaitable d'un historique se situe aux alentours de 4/5 ans (surtout en ce qui concerne le calcul de coefficients saisonniers significatifs).
Cette dimension permet d'atteindre un niveau élevé de signification statistique, elle représente un échantillon suffisant pour qu'il ne s'éloigne pas trop de la population totale qu'il est censé représenter et offre une cohérence économique satisfaisante.

Le nombre d'observations composant un historique est fonction de la périodicité des données. Ainsi, pour une prévision hebdomadaire à un horizon de 15 semaines, il faut disposer de 156 observations, soit 3 ans.

Il ne sert à rien de vouloir remonter trop loin dans le temps ; en effet, des modifications de structure de la chronique (telles que la tendance ou la saisonnalité, ou la corrélation avec des indicateurs) apparaissent avec le temps et risquent de venir perturber la prévision à court terme, voire même de la dégrader.

Dans un cadre de périodicité différent (bimestre, trimestre, quadrimestre ou annuel), nous pourrions retenir moins de réalisations car une valeur plus agrégée représente une quantité d'informations supérieure. En d'autres termes, une donnée annuelle est évidemment moins fine que les douze données mensuelles qui la composent, mais elle contient néanmoins une quantité d'informations, disponible pour l'analyse, infiniment supérieure à n'importe quel chiffre mensuel.

C'est ainsi que des modèles économétriques élaborés à partir de statistiques annuelles sont couramment estimés sur une dizaine d'années, alors qu'il est illusoire d'estimer un modèle de prévision mensuelle sur dix mois.

2. Le calcul des statistiques de base et la présentation des données

2.1. Le calcul des statistiques de base

Dans ce paragraphe, nous présentons les modes de calcul des statistiques simples : caractéristique de valeur centrale et caractéristique de dispersion.

- La moyenne

La formule générale de la moyenne d'une série chronologique de terme général x_t pour laquelle nous disposons de n observations est :

$$\bar{x} = \frac{\sum_{t=1}^n x_t}{n}$$

Le calcul de la moyenne présente un intérêt limité : deux séries des ventes peuvent avoir une moyenne identique alors que leur allure générale est très différente. Egalement dans le cas de séries ayant une forte tendance soit à la hausse soit à la baisse, la moyenne variera fortement à chacune des nouvelles réalisations. C'est pourquoi il est préférable de présenter avec la moyenne un coefficient permettant de prendre en compte la dispersion.

- La dispersion

La variance d'une série chronologique permet d'évaluer la dispersion autour de la moyenne. Nous devons distinguer deux cas, selon que nous sommes en présence de toute la population de façon exhaustive ou que nous disposons seulement d'un échantillon de réalisations considéré comme représentatif de la population totale.

La formule de la variance (*Var*) dans le cas d'une population connue totalement (variance théorique) est :

$$Var(x) = \frac{\sum_{t=1}^{t=n} (x_t - \bar{x})^2}{n}$$

La formule de la variance dans le cas d'un échantillon (variance empirique) est donnée par :

$$Var(x) = \frac{\sum_{t=1}^{t=n} (x_t - \bar{x})^2}{n-1}$$

Dans le cas du calcul de la variance d'une série chronologique, la deuxième formule est la seule retenue ; en effet, nous ne connaissons pas toutes les valeurs, aussi bien les réalisations futures que celles qui sont très éloignées dans le passé ; la série chronologique constitue donc un échantillon représentatif de la série des ventes vraie mais inconnue.

Dans la littérature statistique ou économique, on préfère présenter la racine carrée de la variance appelée écart-type (\mathbf{S}_x) :

$$\mathbf{S}_x = \sqrt{Var(x)}$$

Cet écart-type est l'un des indicateurs fondamentaux de la difficulté à prévoir¹ une chronique, il présente l'avantage de s'exprimer dans la même unité que l'historique. Pour pouvoir interpréter sa valeur, il convient de le rapporter à la moyenne.

- *Coefficient de variation*

Le coefficient de variation se définit comme étant le rapport de l'écart-type à la moyenne calculé à partir de la série brute :

$$CV = \frac{\mathbf{S}}{x}$$

Il rend compte de la difficulté prévisionnelle d'un historique. Plus il est élevé, plus la variance de la série est importante rapportée à la moyenne et, donc, plus la tâche du prévisionniste est *a priori* délicate.

Nous pouvons classer les historiques des ventes en fonction de ce coefficient :

- inférieur à 0,5 : a priori facile à prévoir,
- compris entre 0,5 et 1 : de dispersion moyenne,
- supérieur à 1 : la variance de la série est importante rapportée à sa moyenne et donc, cette série peut s'avérer difficile à prévoir.

Néanmoins le coefficient de dispersion de la série brute ne préjuge pas complètement des difficultés ultérieures : une série fortement fluctuante peut être largement « expliquée » par des coefficients saisonniers très marqués et/ou par un facteur explicatif très influent.

2.2. La présentation des données

Lorsque le prévisionniste dispose d'un historique, le premier réflexe qu'il doit avoir est de tracer le graphique de la série chronologique appelée aussi courbe représentative du phénomène. Cela permet de visualiser l'évolution de la chronique, de déceler les accidents éventuels (pics ou creux), et d'en tirer une tendance générale.

Pour aider à une meilleure visualisation du graphique, il est possible d'appliquer à la série des valeurs brutes une transformation ou un filtre. Une transformation logarithmique² « écrase » les

¹ Nous pouvons noter que si l'écart-type de l'historique est significativement supérieur à celui de l'écart-type de l'erreur de prévision, le système de prévision est performant.

² Attention aux valeurs négatives ou nulles.

valeurs fortes afin de se ramener à une échelle plus restreinte, elle est utilisée pour des séries affectée d'une tendance très forte. Le graphe est alors en échelle semi-logarithmique.

Appliquer un filtre consiste à calculer depuis la série brute une autre série, appelée série filtrée, déduite de la première à partir d'un processus dynamique. Alors que pour une série transformée, il est possible pour chacune de ces valeurs de revenir à la série initiale (par exemple, le passage aux exponentielles pour une chronique préalablement transformée par les logarithmes). Pour une série filtrée, il s'avère impossible de recalculer la série initiale, même en disposant du type de filtre employé.

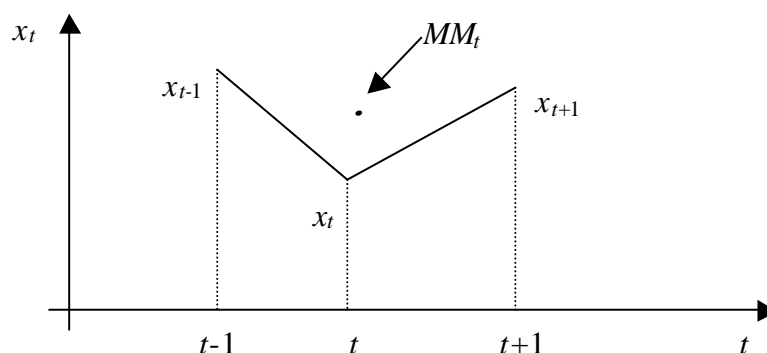
Le filtre le plus employé est la moyenne mobile qui permet d'écarter les pics et les creux afin de mieux déceler la tendance de fond de la chronique. Il s'agit de calculer une moyenne sur un certain nombre de valeurs de la série et de l'affecter à un mois donné.

Soit x_t le terme général de la série brute, la réalisation à la période t , par exemple la vente observée en octobre 2000. La série filtrée MM_t à l'aide d'une moyenne mobile, par exemple, d'ordre 3 a pour expression : $MM_t = (x_{t-1} + x_t + x_{t+1}) / 3$

Il est à noter que cette moyenne mobile d'ordre trois ne peut être calculée que pour $t = 2$ à $n - 1$ (n étant le nombre d'observations), car on ne dispose pas d'observations précédant x_1 et suivant x_n .

Graphiquement, nous observons que la moyenne mobile se trouve au barycentre des trois observations (cf. Graphique 1).

Graphique 1 – Moyenne mobile d'ordre 3



Le choix de l'ordre de la moyenne mobile dépend de l'objectif recherché. Plus l'ordre est élevé, plus la série est lissée et les phénomènes de court terme sont alors gommés, au risque de perdre en réactivité ; seul reste le mouvement de fond (la tendance). L'inconvénient réside alors dans la perte plus grande d'informations en début et fin d'historique. Une mention particulière concerne la moyenne mobile d'ordre douze qui, pour des séries historiques de périodicité mensuelle, représente l'évolution des consommations (ou des ventes) hors phénomènes saisonniers.

Dans les publications statistiques et les revues d'information économique (*BMS*³, *l'Expansion*, *Le Nouvel Economiste*, ...), la plupart des graphiques représentent des phénomènes filtrés par moyenne mobile sur 3 mois. Ceci a pour effet de rendre la courbe plus lisse, de «gommer» les fluctuations en répartissant d'un mois sur l'autre les valeurs extrêmes ; d'où le terme de lissage des données ou de courbe lissée lorsqu'on procède à un filtrage par moyenne mobile.

Les formules générales de filtrage par moyenne mobile sont les suivantes :

³ Bulletin Mensuel de Statistique de l'INSEE.

– Si l'ordre correspond à un nombre impair ($2m + 1$),

$$MM_t = \frac{1}{2m + 1} \sum_{i=-m}^{i=m} x_{t+i}$$

– Si l'ordre correspond à un nombre pair ($2m$), il nous faut recourir à un artifice de calcul afin de faire correspondre le terme central x_t à la valeur de la moyenne mobile MM_t .

$$MM_t = \frac{1}{2m} \left[\frac{1}{2} x_{t-m} + \sum_{i=-m+1}^{i=m-1} x_{t+i} + \frac{1}{2} x_{t+m} \right]$$

Nous voyons que, pour des raisons de commodité des calculs, il est préférable de retenir des moyennes mobiles d'ordre impair⁴.

Le Tableau 2 présente un exemple de calcul d'une moyenne mobile d'ordre trois et d'une moyenne mobile d'ordre six, ainsi que la moyenne, l'écart-type et le coefficient de variation.

Tableau 2 – Exemple de calcul de moyennes mobiles et de coefficients de variation

temps	Consommation	<i>MM3</i>	<i>MM6</i>
1	100,00		
2	89,00	96,00	
3	99,00	81,33	
4	56,00	100,67	92,67
5	147,00	96,67	93,25
6	87,00	96,67	101,50
7	56,00	94,33	113,83
8	140,00	114,33	118,17
9	147,00	147,67	118,42
10	156,00	134,00	126,33
11	99,00	131,00	139,25
12	138,00	112,33	149,25
13	100,00	163,00	149,25
14	251,00	169,00	148,42
15	156,00	184,67	148,00
16	147,00	133,67	149,92
17	98,00	126,33	146,67
18	134,00	119,67	
19	127,00	148,67	
20	185,00		
Moyenne	125,60	125,00	128,21
Ecart-type	44,11	28,28	21,09
CV	0,35	0,23	0,16

Exemple de calculs :

Pour $t = 10$ (moyennes mobiles d'ordre 3 et 6 à la dixième période)

$$MM3_{10} = (147 + 156 + 99) / 3 = 134$$

⁴ La technique des moyennes mobiles est très utilisée par les chartistes afin d'analyser l'évolution des marchés financiers (cf. Béchu et Bertrand, 1997). Dans ce cas particulier, la valeur de la moyenne mobile est indiquée en face de la dernière observation sans se préoccuper d'un quelconque centrage.

$$MM6_{10} = (56/2 + 140 + 147 + 156 + 99 + 138 + 100/2) / 6 = 126,33$$

Nous observons que le coefficient de variation de la série brute est supérieur à celui de la série lissée *MM3* qui lui-même est supérieur à celui de la série *MM6* ; nous retrouvons bien la mécanique du lissage qui écrête les pics et les creux.

3. L'analyse de série chronologique et la prévision

3.1. Les étapes de la prévision des ventes

L'étude des phénomènes économiques a distingué depuis longtemps divers types d'évolution, qui peuvent éventuellement se combiner :

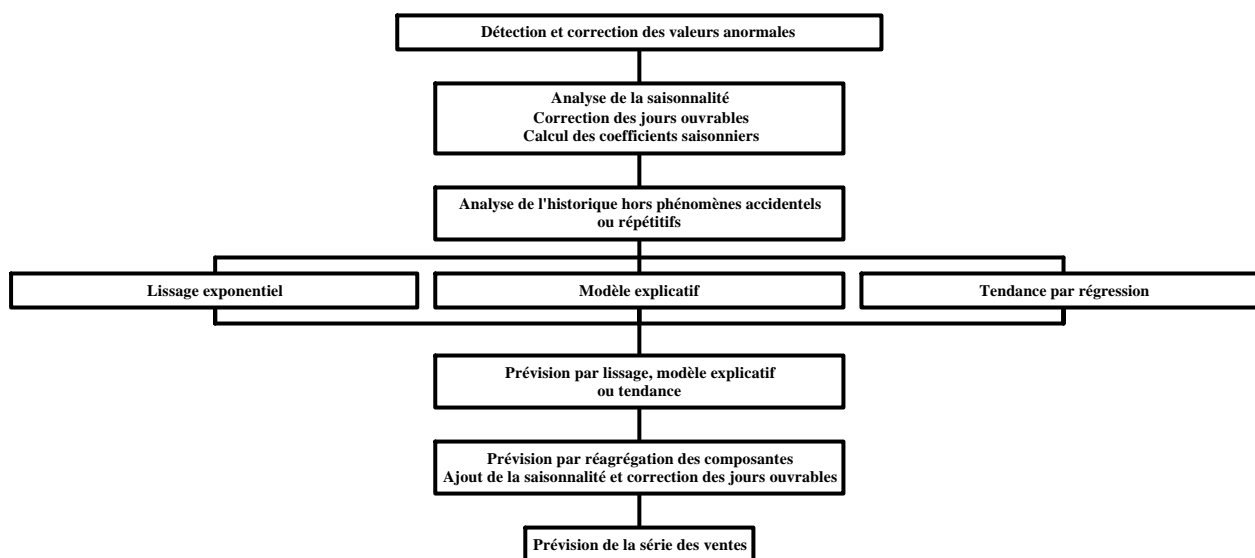
- La tendance (T_t) correspond à une variation lente s'effectuant dans un sens déterminé qui se maintient pendant de longues périodes.
- Le cycle (C_t) est un mouvement d'allure quasi périodique comportant une phase croissante et une phase décroissante. En conjoncture, il est représenté par le cycle de Kitchin d'une période de 4 à 5 ans. Dans la plupart des travaux sur les séries temporelles, la tendance et le cycle sont regroupés en une seule composante.
- La composante saisonnière (S_t) correspond à des variations s'effectuant régulièrement au cours de la semaine, du mois, du trimestre, etc. Elle tient aux saisons, à des habitudes sociologiques et aux rythmes de l'activité humaine.
- Les fluctuations accidentelles (R_t) sont des mouvements erratiques, de fréquences élevées, présentant une allure générale plus ou moins stable. Elles résultent des influences, que toutes sortes d'événements exercent sur la grandeur en cause, si nombreuses que l'on a renoncé à (ou que l'on ne peut) les étudier toutes dans le détail.

Certaines séries statistiques présentent l'un ou l'autre de ces mouvements à l'état pur. Mais la plupart d'entre elles ont une allure plus complexe. Il est possible d'y déceler par exemple une croissance générale et certaines variations saisonnières auxquelles se superposent des fluctuations accidentelles. Remarquons que ces différentes composantes s'entendent pour des séries économiques liées à la conjoncture, le plus souvent mensuelles ou trimestrielles. Dans le domaine de l'entreprise, les composantes sont conservées mais les périodicités sont parfois différentes (hebdomadaire, par exemple).

L'objectif de toute méthodologie prévisionnelle est de décomposer les ventes, que l'on cherche à prévoir, en composantes fondamentales afin de les isoler, puis de les extrapoler indépendamment les unes des autres. La prévision finale est obtenue en réagrégeant tous ces phénomènes.

Le Schéma 1 illustre les étapes classiques de la prévision des ventes.

Schéma 1 – Les étapes de la prévision des ventes



3.2. Recherche et correction de valeurs anormales

Préalablement à tout traitement statistique sur un historique de ventes, il convient d'effectuer une partition entre :

- une activité régulière et habituelle, qualifiée par l'homme d'expérience de normale ;
- un événement accidentel, imprévisible, qualifié alors d'anormal.

Ces « consommations anormales » sont observées fréquemment dans la grande consommation, mais aussi dans d'autres secteurs. Les demandes exceptionnelles provenant de l'exportation ou d'importantes centrales d'achat, ou encore des aléas de la distribution (grèves des transports par exemple) peuvent perturber fortement le processus habituel.

Il s'agit d'un phénomène non répétitif et qui revêt un caractère exceptionnel : par exemple, une vente promotionnelle n'est pas anormale car elle fait partie de la « vie normale » de l'entreprise.

Le problème est donc double : d'une part, identifier l'observation jugée anormale et, d'autre part, la corriger afin de neutraliser ses effets.

3.2.1 La détection des valeurs anormales

Plusieurs méthodes existent sans toutefois qu'aucune ne donne entière satisfaction. Nous présentons succinctement quatre techniques, le lecteur intéressé par les développements mathématiques et une description de la littérature dans le domaine se reportera à la référence citée.⁵

– La méthode de l'intervalle de confiance

Cette méthode suppose la normalité des observations⁶ ; sa mise en œuvre est très simple. Elle consiste à calculer l'écart-type de l'historique des consommations, puis à filtrer toutes les observations qui sortent d'un intervalle de confiance déterminé.

⁵ Vallin (1984).

⁶ L'hypothèse que les observations suivent une loi normale (la fameuse courbe en « cloche ») entraîne une symétrie de la distribution par rapport à la moyenne.

Soit s_x l'écart-type de la série, nous pouvons calculer l'intervalle de confiance de la manière suivante :

$$IC = \bar{x} \pm 1,96 \times s_x$$

où 1,96 est la valeur de la loi normale pour un seuil de confiance de 0,05.

Toute observation extérieure à cet intervalle est considérée comme anormale.

Cette méthode pose des problèmes d'application lorsque les séries sont très saisonnières. En effet, les mois forts ou faibles risquent de « sortir » de l'intervalle de confiance et donc d'être détectés.

En reprenant les données du Tableau 2, l'intervalle de confiance est donné par :

$$IC = 125,60 \pm 1,96 \times 44,11 \text{ soit } IC = [39,14 ; 212,05]$$

L'observation de la période $t = 14$ ($x_{14} = 251$) est extérieure à l'intervalle de confiance, elle est donc détectée comme douteuse.

– *La méthode du test de moyenne*

Cette méthode suppose, là encore, la normalité des observations. Pour tester la conformité de l'observation x_t , on applique un test de comparaison de moyennes entre deux échantillons :

- l'un réduit à un seul élément, l'observation x_t que l'on désire tester ;
- l'autre est composé des $n - 1$ éléments de l'historique (n étant le nombre d'observations).

Nous calculons la statistique suivante (dite du Student empirique) :

$$t_{cal} = \frac{|x_t - \bar{x}|}{\sqrt{\frac{\sum_{j \neq t} (x_j - \bar{x})^2}{(n-1)(n-2)}}$$

Si $t_{cal} > t_{lu}$ (valeur de référence lue dans une table de Student pour $n - 2$ degrés de liberté et un risque α fixé à 0,05), alors l'observation est considérée comme anormale ; dans le cas contraire ($t_{cal} < t_{lu}$), l'observation est classée comme normale.

En reprenant les données du Tableau 2 et en appliquant cette formule à l'observation $t = 14$, nous trouvons les éléments suivants :

$$\bar{x} = 119 \text{ (hors l'observation } x_{14}) ; \sum_{j \neq t} (x_j - \bar{x})^2 = 22\,362 ; t_{cal} = \frac{132}{\sqrt{\frac{22362}{19 \times 18}}} = 16,32$$

$$t_{cal} = 16,32 > t_{lu} = 2,101 \text{ (} \alpha = 0,05 \text{ et pour un nombre de degrés de liberté de } 20 - 2 = 18)$$

Le résultat de ce test corrobore le précédent, l'observation $t = 14$ est anormale.

– *La méthode du double intervalle de confiance*

Cette méthode est une extension de la première et est utilisée pour des séries dont la saisonnalité est très marquée. Elle consiste à calculer un intervalle de confiance pour chaque année et pour chaque mois (ou trimestre). Une observation est considérée comme douteuse si elle « sort » à la fois de l'intervalle de confiance année et de l'intervalle de confiance mois (trimestre). Le Tableau 3 illustre cette méthode.

Tableau 3 – Calcul d'un double intervalle de confiance

	T1	T2	T3	T4	MOYENNE	ECART-TYPE	IC1	IC2
Année 1	164	198	85	179	156,5	43,00	72,21	240,79
Année 2	168	201	98	197	166	41,27	85,10	246,90
Année 3	197	209	100	216	180,5	46,97	88,44	272,56
Année 4	223	245	119	260	211,75	55,14	103,67	319,83
Année 5	298	309	124	267	249,5	74,08	104,31	394,69
MOYENNE	210	232,4	105,2	223,8				
ECART-TYPE	48,91	41,82	14,36	34,53				
IC1	114,13	150,44	77,06	156,11				
IC2	305,87	314,36	133,34	291,49				

IC1 et IC2 sont les bornes de l'intervalle de confiance.

Chaque observation est comparée à l'intervalle de confiance de l'année et du trimestre à laquelle elle appartient. Si elle est extérieure à au moins deux intervalles, l'observation est classée comme étant anormale. Par exemple, la valeur du deuxième trimestre 1989 (Année 3) est comparée à l'intervalle de confiance de son trimestre [150,44 ; 314,36] et de son année [88,44 ; 272,56]. Elle appartient aux deux intervalles de confiance. Nous pouvons conclure qu'elle est donc normale.

Cette dernière méthode fonctionne d'autant mieux que le nombre d'observations est important. Elle est donc particulièrement adaptée aux séries de périodicité mensuelle.

Ces trois méthodes supposent que les observations suivent une loi de Gauss. Or, les séries de ventes obéissent rarement à une telle loi qui impose la symétrie de la répartition par rapport à la moyenne. Cependant, malgré cette limite, ces tests sont les plus utilisés car ils ont un caractère opérationnel simple.

– Détection par régression

Le principe consiste à constituer une variable muette⁷ relative à l'observation que l'on désire tester. Cette variable est composée de 0 sauf pour la période de l'observation potentiellement anormale, pour laquelle on indique 1. Si le coefficient estimé par régression de cette variable est significativement différent de 0, alors l'observation est anormale.

Cette méthode, la plus rigoureuse sur le plan statistique (aucune hypothèse sur la distribution des observations), est efficace pour tester ponctuellement une observation, mais reste lourde à mettre en œuvre dans le cas d'un traitement « industriel », c'est-à-dire automatisé, portant sur de très nombreuses séries des ventes.

3.2.2 La correction des valeurs anormales

L'objectif est de neutraliser les effets de l'observation anormale en la remplaçant par une autre. Nous allons donc préciser les différentes procédures permettant de déterminer cette nouvelle valeur.

⁷ Nous revenons plus en détails sur cette technique qui suppose de connaître le modèle de régression multiple et son interprétation (cf. chapitres 4 et 6).

– *Intervention humaine*

Le plus simple et le plus logique consiste à laisser au gestionnaire la responsabilité de la modification ; en effet, c'est lui – par sa connaissance des causes – qui est le mieux placé pour connaître la part d'anormalité qui existe dans une observation. Cependant, en cas de références logistiques nombreuses (plus d'une centaine), le gestionnaire du système se trouve dans l'incapacité matérielle de consacrer le temps nécessaire à cet exercice.

– *Intervalle de confiance*

L'observation détectée comme anormale est ramenée à la limite haute ou basse de l'intervalle de confiance. Cette méthode est simple dans son fonctionnement et facilement automatisable. Cependant il ne s'agit pas d'une véritable neutralisation car les bornes de l'intervalle sont une limite théorique qui n'a aucune existence et peut donc venir perturber par la suite le calcul des coefficients saisonniers.

– *Valeur estimée par prévision*

Cette méthode consiste à remplacer l'observation anormale en t par la valeur prévue en $t - 1$, à l'aide du modèle de prévision, pour cette même période. Il s'agit d'effectuer une simulation de prévision et non pas de remplacer l'observation détectée par sa valeur ajustée car celle-ci incorpore une partie de l'information défectueuse.

L'avantage de cette méthode réside dans la neutralisation totale de l'observation puisqu'elle est, par construction, conforme au modèle, mais sa contrepartie est une lourdeur informatique accrue.

3.3. Le schéma de décomposition

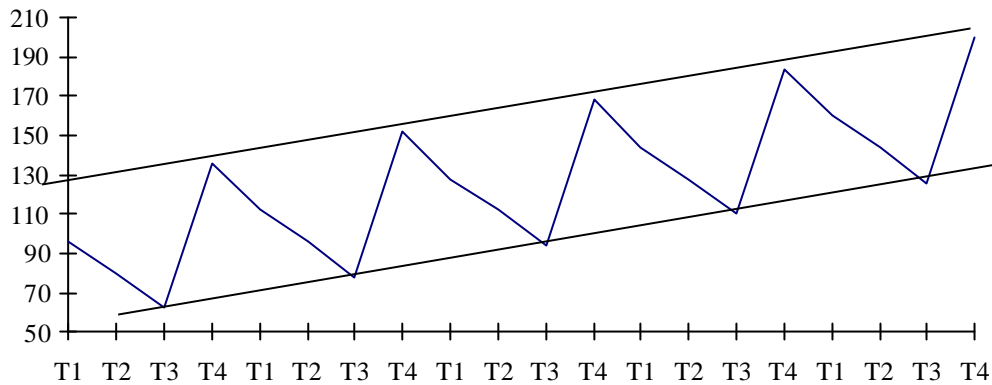
La technique de décomposition–composition repose, bien évidemment, sur un modèle qui l'autorise. Ce modèle porte le nom de schéma de décomposition. Il en existe essentiellement trois grands types :

– Le schéma additif qui suppose l'orthogonalité (indépendance) des différentes composantes. Il s'écrit : $x_t = T_t + S_t + R_t$. Dans ce schéma, la saisonnalité est rigide en amplitude et en période.

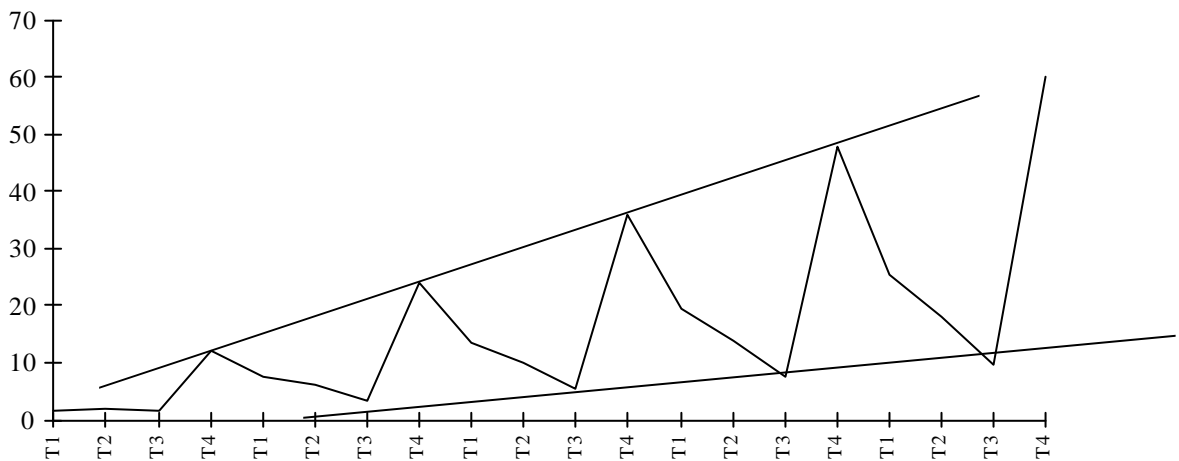
– Le schéma multiplicatif : $x_t = T_t \times S_t + R_t$, dans lequel la composante saisonnière est liée à la tendance (saisonnalité souple avec variation de l'amplitude au cours du temps).

– Le schéma multiplicatif complet : $x_t = T_t \times S_t \times R_t$ (interaction générale des trois composantes). Il est actuellement le plus utilisé en économie. Il est commode puisque le logarithme de la chronique conduit au schéma additif.

Graphique 2 – Exemple de schéma additif



Graphique 3 – Exemple de schéma multiplicatif



Le schéma additif (cf. Graphique 2) correspond à une chronique dont le profil saisonnier ne se modifie pas au cours du temps et dont la tendance est peu marquée.

Le schéma multiplicatif (cf. Graphique 3) rend compte de séries chronologiques dont l'amplitude du profil saisonnier évolue au cours du temps proportionnellement à la valeur de la tendance. Ce schéma se ramène au précédent par passage au logarithme ($\text{Log}(ab) = \text{Log } a + \text{Log } b$)

$$\text{Log}(x_t) = \text{Log}(T_t \times S_t \times C_t \times R_t) = \text{Log } T_t + \text{Log } S_t + \text{Log } C_t + \text{Log } R_t$$

Nous allons étudier successivement ces quatre composantes fondamentales en insistant particulièrement sur les deux plus importantes : tendance et saisonnalité.

3.4. La tendance

Lorsque nous observons un graphe de série chronologique, mentalement nous déduisons une tendance. Notre but ici est de donner des méthodes de calcul simple permettant de connaître avec le maximum de précisions la tendance à moyen terme de la chronique. Pour estimer la tendance, plusieurs procédés élémentaires viennent à l'esprit. En faisant par exemple le rapport des 2 valeurs extrêmes, un pourcentage total d'augmentation est calculé, ce qui nous permet de définir un accroissement mensuel moyen. Le choix des observations extrêmes est arbitraire. Il ne s'impose pas : d'autres observations pourraient être sélectionnées. De plus, on est certain de ne pas tirer le meilleur parti de l'historique, puisqu'un grand nombre de valeurs sont ignorées, en ne prenant que deux observations.

Afin de mieux estimer la tendance, il est nécessaire de faire appel à la statistique mathématique qui fournit un certain nombre d'outils de calcul.

– Tendance par moyenne mobile

Nous avons défini précédemment ce qu'est une moyenne mobile et son effet de lissage sur les courbes. En retenant un ordre de moyenne mobile élevé, nous réduirons les fluctuations de la chronique (plus l'ordre est élevé, plus les fluctuations de la courbe lissée sont amorties).

Les moyennes mobiles présentent deux défauts en tant que technique pour la détermination de la tendance. En premier lieu, elles conduisent à une série plus courte que la série traitée. La tendance n'est obtenue que pour la période allant de la $(m + 1)$ -ième observation à la $(n - m)$ -ième.

En second lieu, aucun principe n'apparaît naturellement pour l'extrapolation d'une tendance déterminée par moyenne mobile. Si l'on veut faire des projections, il faut introduire un modèle plus explicatif pour la représentation de la tendance.

Aussi les moyennes mobiles sont-elles employées soit dans des travaux purement descriptifs d'histoire économique, soit lorsqu'il s'agit d'éliminer (c'est-à-dire soustraire terme à terme) la tendance en vue de l'étude de telle ou telle particularité de la composante résiduelle, par exemple en vue de l'étude des cycles de l'activité.

– Tendance par régression

Le but est d'estimer une tendance à l'aide d'un polynôme :

$$x_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \dots$$

En retenant un polynôme de degré 1 : $x_t = a_1 t + a_0$, on considère que la tendance est linéaire (Graphique 4) et peut être représentée à l'aide d'une droite correspondant à l'ajustement des moindres carrés.

Le but est d'estimer la tendance à long terme de la chronique à l'aide d'une droite (ou par un polynôme de degré plus élevé) estimée par un ajustement des moindres carrés. La technique des Moindres Carrés Ordinaires (MCO) fournit non seulement les estimateurs de a_1 et a_0 respectivement notés \hat{a}_1 et \hat{a}_0 , mais aussi les écarts-types des coefficients et permet ainsi d'effectuer les tests associés (tests des paramètres cf. chapitre 3). Cette méthode postule des paramètres a_1 et a_0 constants dans le temps et la prévision porte le nom d'extrapolation de la chronique. Elle reste utile pour estimer les tendances « lourdes » des phénomènes économiques, « toutes choses étant égales par ailleurs ».

Les coefficients de ce polynôme sont estimés à l'aide des formules suivantes :

$$\hat{a}_1 = \frac{\sum_{t=1}^{t=n} (x_t - \bar{x})(t - \bar{t})}{\sum_{t=1}^{t=n} (t - \bar{t})^2}$$
$$\hat{a}_0 = \bar{x} - a_1 \bar{t}$$

n = nombre d'observations.

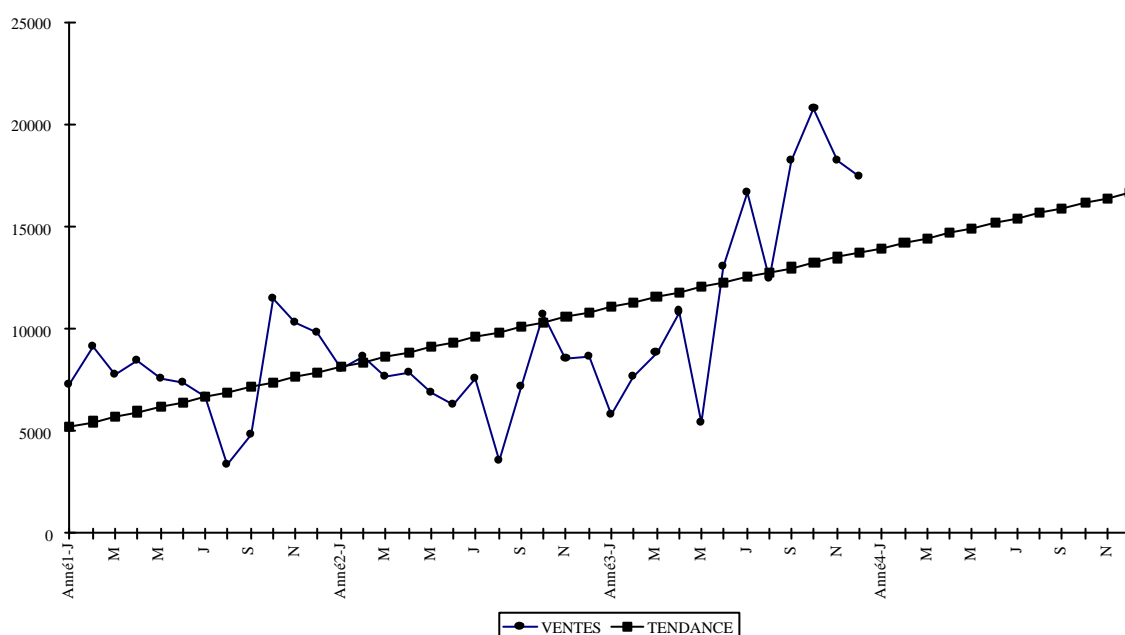
La prévision calculée en n (dernière observation) à l'horizon h est donnée par :

$$x_{n+h} = \hat{a}_1 (n + h) + \hat{a}_0$$

h = horizon de prévision

Les tableurs fournissent directement, par l'appel de fonctions statistiques, les estimations de cette droite de tendance. *Pour un exemple d'utilisation cf. Fichier Web.*

Graphique 4 – Exemple d'ajustement par une droite de tendance



L'équation de la droite ajustée est : $T_t = 243,39 t + 4986,32$

Par rapport à une valeur d'origine de 4986,32, les ventes augmentent – en moyenne sur la période – de 243,39 unités par mois. La prévision consiste donc à extrapoler cette tendance d'évolution.

Le choix du degré de polynôme dépend de plusieurs critères, dont l'horizon de prévision, les ruptures de tendance dans la « vie » du produit et l'historique disponible. Les fonctions du tableur (insertion d'une courbe de tendance) permettent d'estimer différentes formes de fonctions.

L'avantage de la méthode d'estimation de tendance par régression réside dans sa facilité et sa souplesse d'utilisation (choix du degré de polynôme). Cependant ce modèle connaît deux limites fortes :

- il utilise le temps comme unique variable de synthèse,
- les paramètres estimés sont fixes sur la totalité de la période.

3.5. La saisonnalité

3.5.1 Pourquoi analyser une saisonnalité ?

Le problème est maintenant d'interpréter une série éventuellement affectée d'un mouvement saisonnier. On veut apprécier l'évolution de cette série en faisant abstraction de ce mouvement.

Avant d'examiner les méthodes d'étude de la saisonnalité, il faut comprendre les inconvénients que présente pour l'analyse conjoncturelle une méthode souvent employée en pratique, celle qui consiste à comparer le dernier mois (ou trimestre, etc.) observé au mois (ou trimestre, etc.) correspondant de l'année précédente ou bien à comparer directement l'évolution des ventes d'un mois sur l'autre. Le but est de déterminer l'évolution récente du phénomène. Mais les résultats risquent d'être relativement erronés pour deux raisons :

- la tendance moyenne entre un mois (ou trimestre, etc.) et le mois (ou trimestre, etc.) correspondant de l'année précédente peut différer très sensiblement de l'évolution récente ;
- une comparaison assise sur deux observations est affectée par les causes accidentelles qui ont pu agir sur le phénomène à ces dates précises. Un jugement correct sur la tendance conjoncturelle doit reposer sur une analyse de la série observée et non seulement sur la considération de deux observations, aussi judicieux que soit leur choix.

La saisonnalité des ventes d'un article cache la véritable évolution des ventes, *les ventes d'une série brute ne sont donc pas interprétables.*

Par ailleurs, il est plus facile de prévoir les ventes hors phénomènes saisonniers :

- la tendance « vraie » peut être calculée,
- l'impact réel des facteurs explicatifs (publicité, promotion ...) peut être mis en évidence.

3.5.2 La correction des jours ouvrables

Avant de calculer les coefficients saisonniers, il convient de pratiquer une correction des jours ouvrables (CJO) afin de tenir compte du nombre de jours effectivement travaillés dans un mois donné. Par exemple, le mois de février comporte en général moins de jours ouvrables que le mois de mars. Nous devons différencier suivant qu'il s'agit de production (très liée aux jours ouvrables), de livraisons (moins liées) ou de commandes (très peu liées aux jours ouvrables).

Dans le secteur industriel, cette influence est assez faible du fait de l'anticipation des entreprises qui livrent par avance leur client. Dans le secteur des biens de grande consommation, la répercussion des jours ouvrables sur les ventes au consommateur est forte du fait de la fermeture des commerces.

Cette correction s'effectue de la manière suivante :

Soit MJO le nombre moyen de jours ouvrables, par mois dans l'année (environ 21).

La série CJO est égale à : $(x_t \times \text{NJO}) / \text{MJO} \times 100$

avec :

x_t observation du mois t pour la série brute,

NJO nombre de jours ouvrables pour le mois t .

3.5.3 Le calcul de la saisonnalité

Il n'existe pas de méthode parfaitement satisfaisante d'estimation des coefficients saisonniers : quelle que soit la méthode retenue, le risque d'incorporer dans la saisonnalité des fluctuations dues à des valeurs erratiques (appelées indifféremment valeurs aberrantes ou valeurs anormales) est toujours présent. Lors du calcul de la saisonnalité, il convient de faire un certain nombre de choix concernant le type de coefficients saisonniers (additifs/multiplicatifs, fixes/glissants) que nous allons présenter. Les coefficients saisonniers sont légèrement différents selon la méthodologie suivie. Pourtant, ils doivent toujours respecter certaines règles.

a) Le principe de la conservation des aires

La moyenne de la série brute doit être égale à la moyenne de la série CVS (Corrigée des Variations Saisonnières). En effet, l'analyse de saisonnalité a pour but une nouvelle répartition du profil intra-annuel de l'historique, sans modifier le niveau atteint en cumul annuel : les moyennes annuelles de la série brute et de la série CVS doivent être identiques. *Ce principe de base est appelé le principe de la conservation des aires.*

b) Choix du schéma de décomposition

Nous avons vu qu'il existe deux schémas de décomposition :

- Additif $x_t = T_t + S_t + R_t$ (dans ce schéma la saisonnalité est rigide en amplitude et en période)
- Multiplicatif $x_t = T_t \times S_t \times R_t$ (interaction générale des trois composantes)

Dans le cas d'un schéma additif, les coefficients sont exprimés dans la même unité que la chronique ; par exemple pour une série de livraison d'acier en millier de tonnes, si le coefficient du mois d'août est de – 125 cela signifie qu'au mois d'août on livre en moyenne 125 milliers de tonnes d'acier en moins. Dans ce schéma d'analyse, les coefficients saisonniers ne tiennent donc pas compte de la tendance de la série.

La somme des coefficients doit être égale à zéro afin de respecter le principe de la conservation des aires.

Dans le cas d'un schéma multiplicatif, les coefficients saisonniers sont des pourcentages (par exemple, si le coefficient du mois d'août est de 0,82 en moyenne on livre 18 % en moins en août). Dans ce schéma d'analyse, les coefficients saisonniers intègrent donc la tendance de la série. C'est le schéma utilisé de manière classique.

La somme des coefficients est égale au nombre de facteurs saisonniers (la moyenne⁸ doit être égale à 1) : dans un cadre mensuel la somme est égale à 12, pour une série trimestrielle la somme est égale à 4 afin de respecter le principe de la conservation des aires.

Le choix du schéma de décomposition dépend de l'historique. Si la tendance est peu marquée, l'un ou l'autre des deux schémas de décomposition restitue des résultats quasiment identiques. En revanche, si la tendance est marquée, il convient de retenir un schéma multiplicatif dont les effets saisonniers s'adaptent automatiquement – ce sont des pourcentages – au profil de l'historique.

Le seul effet pervers à utiliser des coefficients saisonniers multiplicatifs est rencontré lorsque la série des ventes est affectée par une saisonnalité très marquée ; dans certaines industries, les ventes peuvent être nulles certains mois de l'année. La division ou la multiplication par des valeurs très

⁸ C'est, en fait, le produit des coefficients saisonniers qui doit être égal à 1, cette approximation est valable si les coefficients saisonniers ne sont pas trop dispersés.

faibles amplifie le mouvement saisonnier et peut engendrer une série CVS comprenant des valeurs aberrantes.

c) Coefficients fixes ou glissants ?

- Coefficients fixes : les coefficients calculés sont les mêmes quelle que soit l'année d'analyse.
- Coefficients glissants : les coefficients évoluent tous les ans.

Un mouvement saisonnier est répétitif d'une année sur l'autre, et doit se répéter à l'identique. Il nous semble donc impropre de calculer des coefficients différents par année.

Cependant, dans certaines circonstances où une réflexion économique suggère une évolution des comportements, il peut être intéressant d'intégrer une saisonnalité glissante.

En calculant un coefficient pour chaque mois, le risque d'incorporer une partie de « bruit » dans la saisonnalité s'intensifie. En effet, la distinction entre saisonnalité et résidu sera plus difficile à effectuer en l'absence d'une contrainte de rigidité des coefficients saisonniers. Par exemple si, pour des raisons climatiques, une année a été particulièrement propice à la consommation de bière, une saisonnalité glissante répercutera cette saisonnalité l'année suivante sans aucune raison *a priori*, ne connaissant pas la température à l'avance.

Un autre risque mérite d'être souligné : la confusion possible entre la saisonnalité réelle et une saisonnalité fictive créée par l'entreprise. Il s'agit des entreprises effectuant à la même période de chaque année des promotions ou des variations de tarif. Le calcul des coefficients saisonniers attribue à la saisonnalité cette « sur-vente » due à la politique volontariste de l'entreprise. Un problème surgit alors lorsque l'entreprise modifie la date des promotions ... Dans ce cas, l'utilisation des coefficients saisonniers glissants permet d'intégrer plus rapidement cette modification.

Sauf, réflexion marketing motivant une évolution ou modification prévues des habitudes des consommateurs, on utilise des coefficients saisonniers fixes.

d) Méthodes de calcul

Il n'existe pas de méthode parfaitement satisfaisante d'estimation des coefficients saisonniers. Quelle que soit la méthode retenue, le risque d'incorporer dans la saisonnalité des fluctuations dues à des valeurs erratiques ou à des habitudes promotionnelles est toujours présent.

Cependant, le cas le plus classique consiste à calculer des coefficients saisonniers fixes selon un schéma multiplicatif.

Les coefficients saisonniers sont légèrement différents selon la méthodologie suivie. Nous allons présenter deux exemples de calcul à partir du même historique (il s'agit des ventes d'un produit festif, donc fortement saisonnier) en explicitant les différentes phases.

A) Analyse de saisonnalité selon un schéma additif et des coefficients fixes

Les Tableau 4 et Tableau 5 illustrent le calcul des coefficients saisonniers selon un schéma additif et des coefficients fixes.

Tableau 4 – Analyse de saisonnalité : tableau de calcul selon un schéma additif

DATES	VENTES	MM12	ECART	VentesCVS
Année 1-J	401,60			607,17
F	395,70			628,85
M	451,00			605,58
A	427,60			545,92
M	496,80			637,88
J	467,70			560,96
J	352,30	566	-213	498,96
A	182,10	555	-373	535,99
S	522,20	546	-24	529,76
O	687,20	539	148	552,48
N	1080,3	529	552	579,05
D	1391,6	517	874	573,50
Année 2-J	263,90	517	-253	469,47
F	289,90	520	-230	523,05
M	337,00	519	-182	491,58
A	374,00	518	-144	492,32
M	292,70	512	-219	433,78
J	398,60	504	-106	491,86
J	421,70	506	-84	568,36
A	173,80	513	-339	527,69
S	522,10	517	5	529,66
O	642,40	525	117	507,68
N	984,20	538	446	482,95
D	1307,6	550	757	489,50
Année 3-J	393,40	556	-162	598,97
F	316,20	557	-241	549,35
M	428,60	560	-131	583,18
A	467,60	565	-97	585,92
M	501,00	569	-68	642,08
J	487,40	573	-85	580,66
J	463,30			609,96
A	165,90			519,79
S	595,10			602,66
O	698,10			563,38
N	1012,00			510,75
D	1380,00			561,90
MOYENNE	549			549
E-T	325			51
CV	0,59			0,09

Tableau 5 – Analyse de saisonnalité : coefficients provisoires et définitifs

COEF	PROVISOIRE S^p	DEFINITIF
S1	$(-253,31 - 162,17) / 2 = -207,74$	-205,57
S2	$(-229,85 - 240,77) / 2 = -235,31$	-233,15
S3	$(-182,40 - 131,08) / 2 = -156,74$	-154,58
S4	$(-143,53 - 97,45) = -120,49$	-118,32
S5	$(-218,96 - 67,52) = -143,24$	-141,08
S6	$(-105,56 - 85,30) = -95,43$	-93,26
S7	$(-213,30 - 84,35) / 2 = -148,83$	-146,66
S8	$(-373,36 - 338,75) / 2 = -356,05$	-353,89
S9	$(-24,10 + 4,64) / 2 = -9,73$	-7,56
S10	$(147,88 + 117,22) / 2 = 132,55$	134,72
S11	$(551,72 + 446,45) / 2 = 499,08$	501,25
S12	$(874,40 + 754,47) / 2 = 815,94$	818,108
SOMME	-25,99	0,00
MOYENNE	-2,17	0,00

Le Tableau 4 présente les différentes phases de calcul :

Etape 1 : Estimer la tendance par une moyenne mobile⁹ d'ordre 12 (colonne *MM12*).

⁹ Nous pouvons aussi calculer la tendance à l'aide d'une droite de régression sur le temps.

En appliquant la formule présentée en 1.1, nous ne pouvons pas calculer les six premières valeurs de $MM12$ (de janvier à juin de l'année 1), ainsi que celles de la fin de l'historique (juillet à décembre de l'année 3). $MM12$ ne contient plus de composante saisonnière.

Exemple de calcul pour janvier de l'année 2 :

$$MM12_{J\text{-Année}2} = (0,5 \times 352,30 + 182,10 + 522,20 + \dots + 398,6 + 0,5 \times 421,7) / 12 = 517,20$$

Etape 2 : Calculer les écarts à la série observée.

Nous calculons les écarts observés entre la série brute et la moyenne mobile :

$$e_t = x_t - MM12_t$$

Exemple de calcul pour janvier de l'année 2 : $e_{J\text{-Année}2} = 263,9 - 517,21 = -253,31$

Etape 3 : Rassembler les écarts relatifs aux mêmes mois pour les différentes années et calcul de la moyenne.

Dans le Tableau 5, nous calculons dans la première colonne les coefficients saisonniers provisoires. Il s'agit de rassembler les écarts relatifs au même mois et de calculer la moyenne de ces écarts, ces nombres sont des coefficients saisonniers provisoires, car leur somme n'est pas égale à 0.

1 er mois : $e_{13}, e_{25} \Rightarrow S_1^P = (-253,31 - 162,17) / 2 = -207,74$

2 ème mois : $e_{14}, e_{26} \Rightarrow S_2^P = (-229,85 - 240,77) / 2 = -235,31$

...

12 ème mois : $e_{12}, e_{24} \Rightarrow S_{12}^P = (757,47 + 874,40) / 2 = 815,94$

Etape 4 : Calculer $S = S_1^P + S_2^P + \dots + S_{12}^P$

Les coefficients saisonniers sont ensuite normés afin que leur somme soit nulle.

Si $S = 0$, les $S_1^P, S_2^P, \dots, S_{12}^P$ sont les coefficients saisonniers définitifs.

Si $S \neq 0$, les 12 coefficients définitifs sont donnés par :

$$S_1 = S_1^P - S / 12 ; S_2 = S_2^P - S / 12 ; \dots ; S_{12} = S_{12}^P - S / 12$$

Exemple : $S = -25,99$ d'où $S_2 = -235,31 - (-25,99 / 12) = -233,15$

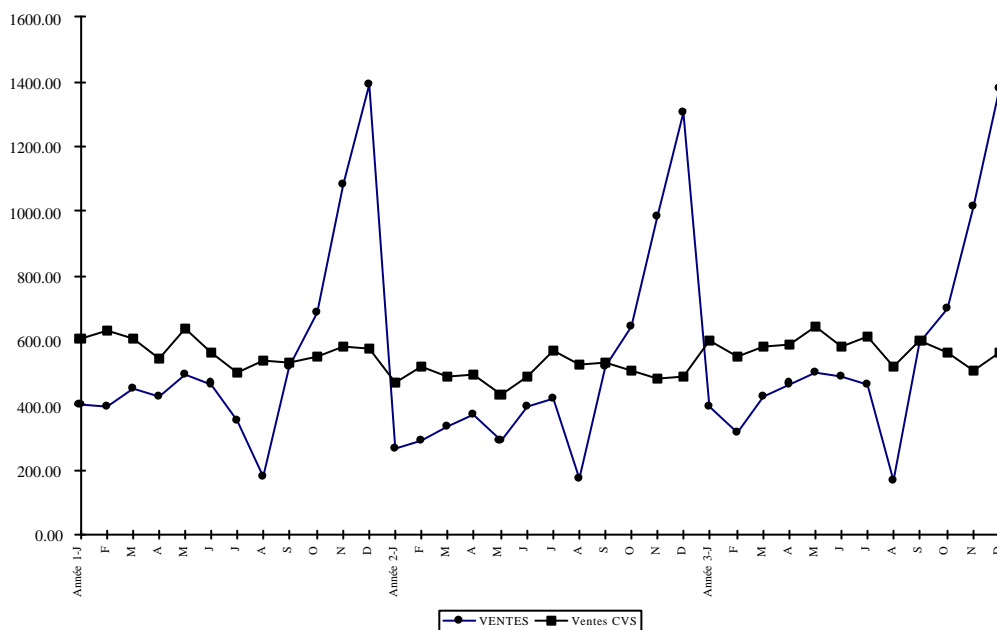
Etape 5 : Calcul de la série CVS

La série CVS est calculée par différence entre la série brute et le coefficient saisonnier du mois considéré.

Pour le mois de janvier de l'année 2, la vente CVS est égale à : $263,90 - (-205,57) = 607,17$

Le Graphique 5 présente la série brute et la série CVS. Nous remarquons que la désaisonnalisation a réduit considérablement la variance de la série (ceci étant illustré par la baisse du coefficient de variation, cf. Tableau 4) et que les ventes, hors saisonnalité, s'avèrent stables et de prévision aisée.

Graphique 5 – Analyse de saisonnalité : coefficients saisonniers fixes et additifs



B) Analyse de saisonnalité selon un schéma multiplicatif et des coefficients glissants

Le Tableau 6 présente le calcul des coefficients saisonniers à partir du même historique que précédemment, selon un schéma multiplicatif et des coefficients glissants.

Tableau 6 – Calcul des coefficients saisonniers glissants et multiplicatifs

DATES	VENTES	MM12	RAPPORT	NORME	Ventes CVS
Année 1-J	401,60		0,51	0,52	768,55
F	395,70		0,56	0,57	692,73
M	451,00		0,65	0,66	678,74
A	427,60		0,72	0,74	577,77
M	496,80		0,57	0,59	847,99
J	467,70		0,79	0,81	577,63
J	352,30	565,60	0,62	0,64	552,28
A	182,10	555,46	0,33	0,34	542,38
S	522,20	546,30	0,96	0,98	533,44
O	687,20	539,32	1,27	1,30	526,62
N	1080,3	528,58	2,04	2,09	516,13
D	1391,6	517,20	2,69	2,76	505,02
Année 2-J	263,90	517,21	0,51	0,54	491,94
F	289,90	519,75	0,56	0,59	494,37
M	337,00	519,40	0,65	0,68	494,03
A	374,00	517,53	0,72	0,76	492,25
M	292,70	511,66	0,57	0,60	486,67
J	398,60	504,16	0,79	0,83	479,53
J	421,70	506,05	0,83	0,88	481,33
A	173,80	512,55	0,34	0,36	487,51
S	522,10	517,46	1,01	1,06	492,18
O	642,40	525,17	1,22	1,29	499,52
N	984,20	537,75	1,83	1,92	511,49
D	1307,6	550,13	2,38	2,50	523,26
Année 3-J	393,40	555,57	0,71	0,70	565,43
F	316,20	556,97	0,57	0,56	566,86
M	428,60	559,68	0,77	0,75	569,62
A	467,60	565,05	0,83	0,81	575,08
M	501,00	568,53	0,88	0,87	578,62
J	487,40	572,71	0,85	0,84	582,88
J	463,30		0,83	0,82	565,85
A	165,90		0,34	0,33	497,94
S	595,10		1,01	0,99	600,28
O	698,10		1,22	1,20	580,85
N	1012,10		1,83	1,80	562,82
D	1380,00		2,38	2,34	590,90
MOYENNE	549,24				558,07
E-T	320,46				79,42
CV	0,58				0,14

Les différentes étapes de calcul sont les suivantes.

Etape 1 : Estimer la tendance par moyenne mobile.

Cette étape est identique à celle du schéma additif.

Etape 2 : Calculer le rapport entre la série observée et la moyenne mobile.

Nous sommes en schéma multiplicatif, nous effectuons donc le rapport entre la série brute et la moyenne mobile.

$$S_t^p = x_t / MM12_t$$

En ce qui concerne les six premiers mois et les six derniers mois, la moyenne mobile d'ordre 12 ne peut pas être calculée. Afin de compléter cette colonne, on intègre la valeur du même mois de l'année suivante pour les six premiers mois et la valeur du même mois de l'année précédente pour les six derniers mois.

Etape 3 : Normer les coefficients (conservation des aires).

Les coefficients saisonniers sont ensuite normés afin que leur moyenne, pour une année donnée, soit égale à 1.

Nous normons les coefficients saisonniers année par année afin que leur somme soit égale à 12.

$$\text{Calcul de } S = S_1^P + S_2^P + \dots + S_{12}^P$$

$$S_1 = (S_1^P \times 12 / S) ; S_2 = (S_2^P \times 12 / S) ; \dots ; S_{12} = (S_{12}^P \times 12 / S).$$

Etape 4 : Calcul de la série CVS

La série CVS est calculée par le rapport entre la série brute et le coefficient saisonnier du mois considéré.

Exemple de calcul pour février de l'année 2.

Calcul de la somme des 12 coefficients provisoires de janvier à décembre de l'année 2 :

$$S = 0,51 + 0,56 + 0,65 + \dots + 2,38 = 11,41$$

$$S_{\text{fev-Année 2}} = (0,56 \times 12) / 11,41 = 0,59$$

Pour le mois de février de l'année 2, la vente CVS est égale à : $289,90 / 0,59 = 494,37$

C) Analyse de saisonnalité selon un schéma multiplicatif et des coefficients fixes (Fichier Web).
Le lecteur analysera lui-même les résultats à partir du fichier Excel.

e) Autre méthode de désaisonnalisation : CENSUS

En 1954, Shiskin, spécialiste des cycles économiques, propose une méthode de désaisonnalisation utilisant, de façon itérative, plusieurs moyennes mobiles. Shiskin, chercheur au « Bureau of Census » des Etats-Unis, a donné le nom de Census-1 à cette méthode. Depuis 1954, elle a connu de nombreuses améliorations : Census-2, Census-3 et en 1967 Shiskin associé à Young et Musgrave a proposé la version Census $\times 11$ qui est, actuellement, la méthode la plus utilisée pour désaisonnaliser les chroniques (notamment par l'INSEE). Il convient de noter l'amélioration apportée à cette technique en 1979 par Dagum qui propose la version Census 11 *ARMMI*.

CENSUS II permet une analyse de saisonnalité par des moyennes mobiles en particulier à l'aide des formules de Spencer (moyenne mobile d'ordre 15). Il est largement diffusé et son emploi s'est généralisé dans beaucoup d'organismes d'études statistiques mais peu en entreprises. Dans la méthode CENSUS II, des coefficients saisonniers glissants sont calculés en vue d'intégrer des modifications de structure de la saisonnalité. Ceci est donc critiquable comme nous l'avons souligné.

Comme la méthode Census est bâtie à partir d'itérations successives de moyennes mobiles d'ordre différent pour mieux appréhender l'extra saisonnalité de la chronique ainsi que les fluctuations de la saisonnalité, elle fait perdre de l'information à l'extrémité terminale de la chronique. C'est pourquoi la méthode Census ne peut être employée que pour des séries de ventes longues (10 ans), ce qui en limite l'utilisation opérationnelle en entreprise.

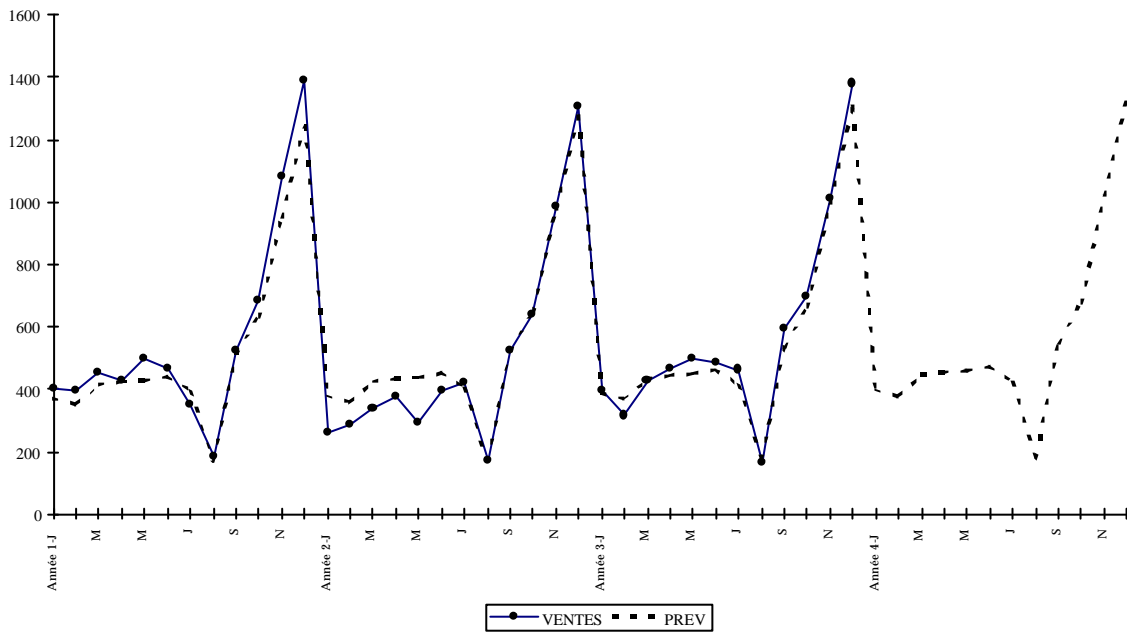
Comme nous pouvons l'entrevoir, l'analyse de saisonnalité reste un traitement relativement complexe qui nécessite rapidement des moyens informatiques. C'est une étape importante de la

démarche prévisionnelle, car les coefficients saisonniers permettent de répartir un niveau annuel escompté à l'intérieur de l'année.

f) Une première prévision par extrapolation de la tendance et de la saisonnalité

Le Graphique 6 illustre une prévision calculée par l'extrapolation de la tendance et de la saisonnalité. Pour les calculs de la tendance et de la saisonnalité déjà présentés, le lecteur se réfère au fichier Excel.

Graphique 6 – Prévision par extrapolation de la tendance et de la saisonnalité



3.5.4 Test sur la saisonnalité

Avant de corriger une chronique de ces variations saisonnières, il convient d'élaborer un test permettant de détecter l'existence d'une saisonnalité. Le test le plus communément employé est celui de Fisher par analyse de la variance du facteur mensuel (ou trimestriel) par rapport à la variance totale de la série.

La question posée est la suivante : l'adjonction d'une saisonnalité à la tendance contribue-t-elle à améliorer de manière significative l'explication de la série brute ? Cela revient à comparer :

- la somme des carrés des écarts – entre prévision et réalisation – du modèle à tendance seule,
- à la somme des carrés des écarts du modèle à tendance et saisonnalité.

Pour répondre à cette question, nous recourons à un test de Fisher exposé ci-après.

a) Calculer la somme des carrés des écarts du modèle avec tendance simple.

Estimation de la droite de tendance : $T_t = \hat{a}_1 t + \hat{a}_0$, puis calcul de :

$$U^* = \sum_{t=1}^n (x_t - T_t)^2$$

Pour U^* , le nombre de degrés de liberté¹⁰ est donc $ddl_{U^*} = n - 2$ ($n =$ nombre total d'observations) car nous avons estimé deux paramètres a_0 et a_1 .

b) Calculer la somme des carrés des écarts du modèle avec tendance et saisonnalité. Estimation des coefficients saisonniers S_t , puis calcul de :

$$U^{**} = \sum_{t=1}^n (x_t - (T_t \times S_t))^2$$

Pour U^{**} , le nombre de degrés de liberté est donc $ddl_{U^{**}} = n - 2 - 11$ car nous avons estimé deux paramètres a_0 et a_1 et 11 coefficients saisonniers (le douzième se déduit des onze autres d'après le principe de la conservation des aires).

c) Calculer la valeur du Fisher empirique.

$$F^* = ((U^* - U^{**})/11) / (U^{**} / (n - 13))$$

La valeur du F^* empirique, calculé à partir de l'historique, est à comparer à la valeur du Fisher théorique donnée par la table de la loi¹¹ de Fisher-Snedecor aux degrés de liberté correspondants ($v_1 = ddl_{U^*} - ddl_{U^{**}} = 11$; $v_2 = ddl_{U^{**}} = n - 13$).

Si $F^* > F$ lu alors la série est saisonnière, sinon la série n'est pas saisonnière.

Dans le cas de chroniques relativement longues (4 ans en données mensuelles), F lu peut être approché par 2 ; cela évite une lecture systématique de la table pour un risque d'erreur très faible. Le Tableau 7 illustre la méthode de calcul.

¹⁰ La notion de degrés de liberté correspond au nombre de valeurs restant réellement à disposition après une procédure d'estimation statistique. Si un échantillon comprend 10 observations et qu'on dispose en plus de la moyenne de cet échantillon, on ne peut choisir librement les valeurs que pour 9 de ces observations, la dixième se déduisant de la valeur de la moyenne. Le nombre de degrés de liberté est donc de $n - 1 = 9$.

¹¹ Les tables statistiques sont en fin d'ouvrage.

Tableau 7 – Test de détection de saisonnalité

DATES	VENTES	Tt	ECART CARRE	Vt/Tt	St	ECART CARRE
Année 1-J	401,60	418	269	0,96	0,79	5330
F	395,70	426	889	0,93	0,76	5174
M	451,00	433	324	1,04	0,85	6722
A	427,60	441	167	0,97	0,79	6042
M	496,80	448	2381	1,11	0,91	8156
J	467,70	456	149	1,03	0,84	7229
J	352,30	463	12256	0,76	0,62	4102
A	182,10	471	83177	0,39	0,32	1096
S	522,20	478	1953	1,09	0,89	9012
O	687,20	486	40682	1,42	1,16	15606
N	1080,3	493	344920	2,19	1,79	38567
D	1391,6	500	794059	2,78	2,27	63996
Année 2-J	263,90	508	59584	0,52	0,58	998
F	289,90	515	50894	0,56	0,63	1205
M	337,00	523	34595	0,64	0,72	1628
A	374,00	530	24491	0,71	0,79	2005
M	292,70	538	60169	0,54	0,61	1228
J	398,60	545	21577	0,73	0,82	2278
J	421,70	553	17237	0,76	0,85	2550
A	173,80	560	149529	0,31	0,35	433
S	522,10	568	2106	0,92	1,03	3908
O	642,40	575	4477	1,12	1,25	5917
N	984,20	583	160973	1,69	1,89	13887
D	1307,6	590	514254	2,21	2,48	24513
Année 3-J	393,40	598	41854	0,66	0,74	2451
F	316,20	605	83684	0,52	0,59	1583
M	428,60	613	33996	0,70	0,79	2909
A	467,60	620	23372	0,75	0,85	3462
M	501,00	628	16123	0,80	0,90	3975
J	487,40	635	21927	0,77	0,86	3762
J	463,30	643	32283	0,72	0,81	3399
A	165,90	650	234812	0,26	0,29	436
S	595,10	658	3953	0,90	1,02	5608
O	698,10	665	1065	1,05	1,18	7717
N	1012,10	673	115009	1,50	1,69	16221
D	1380,00	680	489343	2,03	2,28	30157
SOMME	19773		U* = 3478534			U** = 313262

$$F^* = ((U^* - U^{**})/11) / (U^{**} / (n - 13)) = ((3478534 - 313262)/11) / (313262 / (36 - 13)) = 21,13$$

F lu à 11 et 23 degrés de liberté = 2,24

La série est donc saisonnière (ce qui se distingue à l'examen du Graphique 5).

3.5.5 Famille de saisonnalité

Il s'avère parfois préférable de calculer les coefficients saisonniers non pas à partir de l'historique de l'article lui-même mais sur un agrégat d'articles appelé « famille de saisonnalité ». Cette famille de saisonnalité est donc composée d'articles homogènes vis-à-vis des variations saisonnières, c'est pourquoi elle est souvent différente, en terme de composition d'articles, de la famille commerciale.

Les avantages sont manifestes :

- les coefficients saisonniers sont plus robustes car calculés sur un agrégat,
- les effets sur la saisonnalité d'un « cannibalisme » ponctuel (un article est en promotion au détriment des ventes d'un autre) sont neutralisés,
- un article nouveau avec un historique faible peut être rattaché à une famille de saisonnalité et ainsi la prévision de cet article intègre les variations saisonnières.

3.6. La composante cyclique

La tendance et la composante saisonnière sont souvent les caractéristiques les plus évidentes d'une série chronologique, particulièrement dans le domaine économique. Mais si on prend une chronique pour laquelle ces composantes sont de peu d'importance, ou si on peut mentalement les filtrer et les ignorer (par observation visuelle), la caractéristique la plus visible qui subsiste dans une chronique suffisamment longue est une série de fluctuations irrégulières. Quelquefois ces fluctuations ont une amplitude et une fréquence régulière (il s'agit alors véritablement de cycles), principalement de périodes longues (deux ans ou plus), mais bien souvent il s'agit de simples phénomènes de hausse et de baisse irréguliers.

L'analyse spectrale¹² est la seule méthode spécifique d'analyse de cycles. Son but est de décomposer ces fluctuations en cycles de périodes et d'amplitudes différents.

Les cycles étudiés à un niveau macro-économique (cycles de Kondratieff, de Kuznets, de Juglar, etc.) ne sont pas observés directement au niveau micro-économique où les secteurs sont affectés de cycles de période plus courte.

Les méthodes spectrales, dérivées de la physique (décomposition de Fourier), fournissent une approche mathématique naturelle à ce mélange de régularité et d'irrégularité dans les séries chronologiques.

Leur démarche est de décomposer la série en une somme finie de sinusoi des et de calculer pour chacune d'elles la contribution à la variance générale de la série.

Le plus souvent, cette composante cyclique est ignorée soit qu'il est impossible de la mettre en évidence du fait de son inexistence, soit que nous ne disposons pas d'un historique assez long. La composante cyclique est traitée par de nombreux auteurs comme combinée à la tendance de moyen terme : cette composante mixte est appelée en anglais trend-cycle. Elle peut également être intégrée au résidu (le facteur résiduel explicité ci-dessous), considéré comme une série à prévoir épurée de la tendance et de la saisonnalité. Cela justifie l'emploi ultérieur de méthodes explicatives.

3.7. Le facteur résiduel

En reprenant le schéma de décomposition : $x_t = T_t + S_t + C_t + R_t$, il nous est possible maintenant d'isoler le facteur résiduel : $R_t = x_t - T_t - S_t - C_t$

Cette composante représente les fluctuations erratiques de la chronique dues aux accidents conjoncturels. Il est intéressant de savoir s'il s'agit de phénomènes purement aléatoires au sens gaussien du terme (espérance mathématique nulle, symétrie de la distribution par rapport à la moyenne, constance de la variance), ou s'il est possible de trouver une loi de reproduction du processus. Nous pouvons distinguer les deux termes, résiduel et aléatoire, de la façon suivante :

- facteur résiduel : il est possible de trouver une loi de reproduction de cette composante,
- facteur aléatoire : il n'est pas possible de l'expliquer.

En vue d'établir avec certitude dans lequel des deux cas nous nous situons, le calcul du corrélogramme¹³ (ou fonction d'autocorrélation) fournit une indication sur la nature de cette composante. Nous pourrions employer alors le terme de facteur résiduel ou le terme de facteur aléatoire de façon adaptée (c'est-à-dire un processus pour lequel on ne peut plus rien expliquer).

¹² Cf. Bourbonnais et Terraza (1998) Chapitre 4.

¹³ Le calcul du corrélogramme et son interprétation sont examinés au chapitre suivant.
Régis Bourbonnais - Prévision des ventes

Chapitre 3

PRINCIPALES METHODES (1) :

LES METHODES FONDEES SUR LE LISSAGE EXPONENTIEL

Les développements de la pratique statistique ont permis de disposer d'un certain nombre d'outils de calcul. Ce chapitre a pour finalité de passer en revue les méthodes de prévision extrapolatives fondées sur le principe du lissage exponentiel, en indiquant pour chacune d'elles les développements théoriques minimaux nécessaires à leur compréhension. Le lecteur intéressé par plus de détails méthodologiques se reportera en fin d'ouvrage où une bibliographie lui fournira les références.

Les techniques de lissage exponentiel ont été introduites par Holt en 1957 mais surtout par Brown en 1962. Le lissage regroupe l'ensemble des techniques empiriques qui ont pour caractéristiques communes d'accorder un poids plus important aux valeurs récentes de la chronique.

1. Caractéristiques des méthodes de lissage

1.1. Principes de base

– *Premier principe : la dévalorisation croissante de l'information avec l'âge*

La méthode du lissage exponentiel repose sur l'idée de départ que les informations contenues dans une série chronologique ont d'autant plus d'importance qu'elles sont plus récentes. Pour effectuer une prévision il faut donc affecter aux informations un poids d'autant plus faible qu'elles proviennent d'époques plus éloignées.

– *Deuxième principe : la synthétisation des informations*

L'historique complet d'une série chronologique est difficile à manier.

La technique du lissage exponentiel permet de condenser cet historique sous forme de quelques paramètres. Pour effectuer une prévision à l'aide de cette technique, il n'est nécessaire de conserver – en mémoire – que quelques valeurs.

– *Troisième principe : la réactualisation permanente des paramètres moyennant des calculs relativement simples*

La méthode du lissage exponentiel est adaptative, c'est-à-dire qu'elle reprend en permanence les paramètres, avec la même périodicité que celle qui préside à l'arrivée des informations. Ce principe n'est d'ailleurs qu'une conséquence de l'association des deux principes précédents.

Nous allons démontrer ces trois principes.

1.2. Formulation

Supposons que x_t représente les ventes d'un produit quelconque à la date t . Ces ventes x_t peuvent être considérées comme le résultat d'une combinaison linéaire infinie de ses valeurs passées, le poids (ou l'influence) du passé sur le présent étant décroissant avec son ancienneté.

Notations

x_t = réalisation de la chronique (les ventes) x à la période t , (un mois donné par exemple),

S_t = valeur lissée de la chronique,

\hat{x}_t = prévision de la chronique x_t pour $t + 1$ calculée en t , c'est-à-dire que x_t doit être comparé à \hat{x}_{t-1} ,

α = coefficient de lissage, avec $\alpha \in [0 ; 1]$.

Le principe de base du lissage exponentiel nous propose pour \hat{x}_t :

$$\hat{x}_t = S_t = \hat{x}_{t-1} + \alpha (x_t - \hat{x}_{t-1}) \quad [1]$$

Le lissage apparaît comme le résultat de la dernière valeur lissée corrigé par une pondération de l'écart entre la réalisation et la prévision. Nous retrouvons le troisième principe de l'adaptation du lissage à l'erreur de prévision.

L'équation [1] peut se modifier de la façon suivante

$$\hat{x}_t = \alpha x_t + (1 - \alpha) \hat{x}_{t-1} \quad [2]$$

Sous cette forme, le lissage apparaît comme étant une moyenne pondérée de la dernière réalisation et de la dernière valeur lissée.

Pour observer l'influence du coefficient α , nous posons dans la première équation :

– $\alpha = 0$ alors $\hat{x}_t = \hat{x}_{t-1}$, ce qui signifie que les nouvelles observations ne sont pas intégrées pour calculer les prévisions, le lissage est « inerte » par rapport aux réalisations, les prévisions restent inchangées.

– $\alpha = 1$ alors $\hat{x}_t = x_t$

Ici, le modèle suit les dernières informations, la nouvelle valeur lissée est toujours égale à la dernière réalisation, le lissage est hyper-réactif.

La relation [2] peut être développée en remontant dans le temps ($t - 1, t - 2, \dots, t - n, \dots, 0$) et laisse apparaître que la nouvelle valeur lissée \hat{x}_t est une combinaison linéaire de toutes les observations du passé, affectées d'un poids décroissant avec l'âge. Les poids sont de plus en plus faibles au fur et à mesure que l'on s'éloigne de l'observation actuelle.

$$\hat{x}_t = \alpha x_t + \alpha (1 - \alpha) x_{t-1} + \alpha (1 - \alpha)^2 x_{t-2} + \alpha (1 - \alpha)^3 x_{t-3} + \dots + \alpha (1 - \alpha)^{n-1} x_{t-n+1} + (1 - \alpha)^n x_0 \quad [3]$$

Nous avons ainsi démontré les premier et deuxième principes :

– Puisque le coefficient α est compris entre 0 et 1, le poids accordé aux valeurs est géométriquement décroissant, comme l'illustrent le Tableau 8 et le Graphique 7.

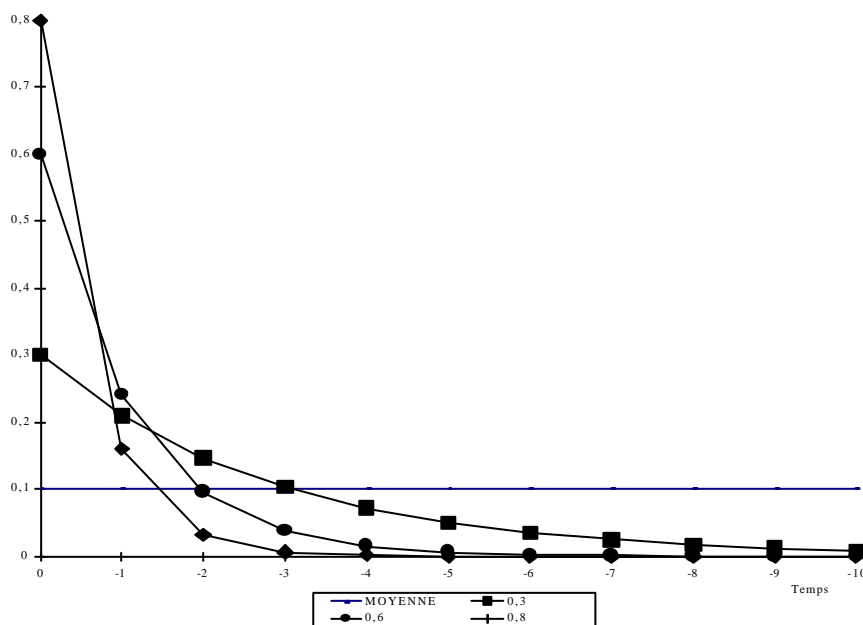
– Dans la dernière valeur lissée est contenue – de manière synthétique – toute l'information historique.

Tableau 8 – Dévalorisation de l'information avec son ancienneté
 Comparaison entre une moyenne classique et trois valeurs de α

Temps	Moyenne classique	Poids	Pondération géométriquement décroissante pour $\alpha = 0,30$
0	0,1	α	0,30
-1	0,1	$\alpha (1 - \alpha)$	0,21
-2	0,1	$\alpha (1 - \alpha)^2$	0,15
-3	0,1	$\alpha (1 - \alpha)^3$	0,10
-4	0,1	$\alpha (1 - \alpha)^4$	0,07
-5	0,1	$\alpha (1 - \alpha)^5$	0,05
-6	0,1	$\alpha (1 - \alpha)^6$	0,04
-7	0,1	$\alpha (1 - \alpha)^7$	0,02
-8	0,1	$\alpha (1 - \alpha)^8$	0,02
-9	0,1	$\alpha (1 - \alpha)^9$	0,01
-10	0,1	$\alpha (1 - \alpha)^{10}$	0,01

Pour $\alpha = 0,30$ la dernière valeur x_t de la chronique est pondérée par 0,3. L'avant-dernière x_{t-1} est pondérée par 0,21 (c'est-à-dire que 21 % de sa valeur contribue au calcul de la prévision) et ainsi de suite ..., à partir de la huitième valeur x_{t-8} le coefficient de pondération est inférieur à 0,02.

Graphique 7 – Dévalorisation de l'information avec son ancienneté
 Comparaison entre une moyenne classique et trois valeurs de α



1.3. Rôle de la constante de lissage

Nous avons observé précédemment les conséquences de deux choix extrêmes : $\alpha = 0$ et $\alpha = 1$. Un compromis doit donc être trouvé entre stabilité, c'est-à-dire effacement des variations purement aléatoires, et rapidité de réponse pour repérer des changements de tendance. Le paramètre α , appelé la constante de lissage, joue un rôle très important :

- lorsque α est proche de 0, la pondération s'étale sur un grand nombre de termes du passé, la mémoire du phénomène étudié est forte et la prévision est peu réactive aux dernières observations ;
- lorsque α est proche de 1, les observations les plus récentes ont un poids prépondérant sur les termes anciens, la mémoire du phénomène est faible et le lissage est très réactif aux dernières observations.

Il est possible de calculer le délai moyen de réaction – certains auteurs parlent d'âge moyen de l'information – qui est la moyenne pondérée des coefficients de lissage. Il est donné par :

$$\bar{D} = \frac{1-\alpha}{\alpha}$$

Si $\alpha = 1$, l'âge moyen est nul puisque seule la dernière valeur est prise en compte ; si $\alpha = 0$ l'âge moyen est infini puisque seule la valeur initiale est prise en compte.

Le lissage joue un rôle de réducteur de variance, on démontre la formule suivante :

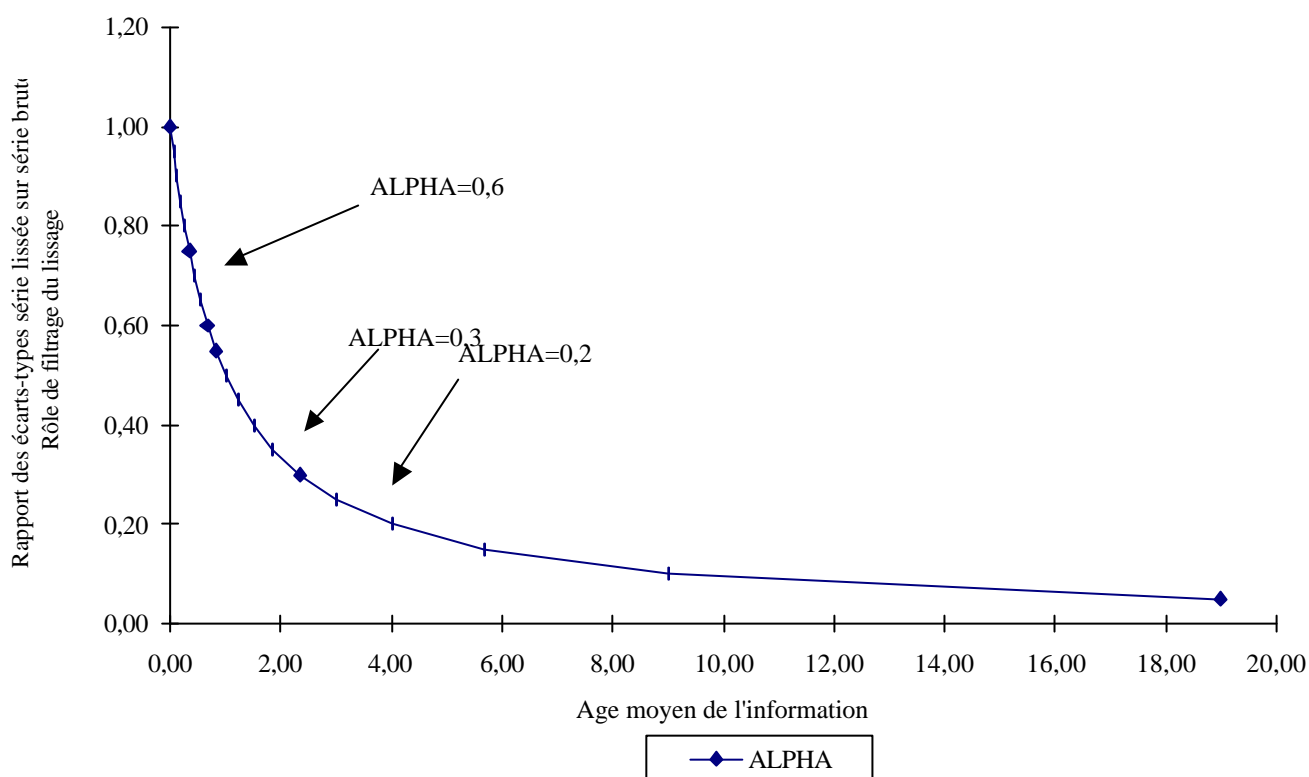
$\frac{Var(\hat{x}_t)}{Var(x_t)} = \frac{\alpha}{2-\alpha}$, si $\alpha = 0$, \hat{x}_t est de variance nulle (puisque la prévision est constante quelles que soient les valeurs de x_t), si $\alpha = 1$, \hat{x}_t a la même variance que x_t . (la nouvelle valeur lissée est toujours égale à la dernière réalisation).

Pour assurer un rôle de filtrage efficace, il convient de choisir un α faible, mais alors on perd en réactivité.

Le Graphique 8 présente en ordonnée les valeurs du rapport de l'écart-type de la série lissée à l'écart-type de la série brute ($\sqrt{\frac{\alpha}{2-\alpha}}$) et, en abscisse, l'âge du lissage pour différentes valeurs de α .

A la lecture de cette courbe, on comprend pourquoi la valeur de $\alpha = 0,3$ est très couramment choisie : lorsque le coefficient α est inférieur à 0,20, l'âge du lissage s'accroît très vite pour un gain faible de filtrage ; au-delà de 0,4, la qualité du filtrage décroît très vite pour une faible réduction de l'âge moyen.

Graphique 8 – Valeur du coefficient α , âge moyen de l'information et filtrage



Nous reviendrons en détail à la section 5 sur le choix optimal de la (ou des) coefficient (s) de lissage.

2. Le lissage simple : le modèle stationnaire

Les formules de lissage simple sont donc les suivantes :

$$\hat{x}_t = S_t = \alpha x_t + (1 - \alpha) \hat{x}_{t-1}$$

avec $\hat{x}_1 = x_1$ pour initialiser

et la prévision calculée en n à l'horizon de h périodes est égale à :

$$\hat{x}_{n+h} = \hat{x}_n \quad \forall h$$

Nous observons que la prévision est constante quel que soit l'horizon.

Le Tableau 9 illustre un calcul.

Tableau 9 – Prédiction à partir du modèle de lissage exponentiel simple

t	x_t	\hat{x}_t	$e_t = x_t - \hat{x}_t$
1	30	30,00	0
2	40	33,00	10,00
3	40	35,10	7,00
4	30	33,57	-5,10
5	20	29,50	-13,57
6	20	26,65	-9,50
7	30	27,65	3,35
8	30	28,36	2,35
9		28,36	
10		28,36	
11		28,36	

On initialise $\hat{x}_1 = x_1 = 30$

Pour $t = 2$ (prédiction calculée en $t = 1$ pour $t = 2$) nous avons :

$$\hat{x}_2 = 0,3 \times x_2 + 0,7 \times \hat{x}_1 = 0,3 \times 40 + 0,7 \times 30 = 33 = x_1$$

Pour $t = 3$: $\hat{x}_3 = 0,3 \times 40 + 0,7 \times 33 = 35,10$

...

Pour $t = 8$, $\hat{x}_8 = 0,3 \times 30 + 0,7 \times 27,65 = 28,36$

Pour $t = 9, 10, 11$ $\hat{x}_{11} = \hat{x}_{10} = \hat{x}_9 = 28,36$

Nous vous proposons à titre d'exercice de charger le fichier Web et d'effectuer la simulation demandée.

La simulation a montré que si la série est affectée d'une tendance, le lissage exponentiel simple est défaillant. Or, en prévision des ventes, la plupart des séries sont affectées d'une tendance à la hausse ou à la baisse. Dans la pratique, le lissage simple n'est pas utilisé. C'est pourquoi on préfère utiliser le lissage exponentiel double qui permet d'intégrer ces phénomènes de tendance.

3. Le lissage exponentiel double : le modèle linéaire

Les formules précédentes permettent de calculer une prédiction pour des séries chronologiques stationnaires, sans tendance. Nous pouvons définir un lissage exponentiel double qui est utilisé en cas de série avec tendance :

Le modèle du lissage exponentiel double (LED) s'applique à une chronique du type :

$$x_t = a_{0t} + a_{1t} t$$

Nous remarquons qu'il s'agit de la même spécification que pour une droite de tendance (cf. chapitre 2), la moyenne (\hat{a}_{0t}) et la tendance (pente de la droite \hat{a}_{1t}) évoluent au cours du temps.

Comme son nom l'indique, la technique du LED consiste à effectuer un lissage de la série déjà lissée.

On démontre¹⁴ les formules suivantes :

$$S_t = \mathbf{a} x_t + (1 - \mathbf{a}) S_{t-1}$$

$$SS_t = \mathbf{a} S_t + (1 - \mathbf{a}) SS_{t-1} \text{ (d'où le terme de lissage double)}$$

¹⁴ Cf. Bourbonnais et Terraza 1998, page 55.

$$\begin{cases} a_{1t} = \frac{\mathbf{a}}{1-\mathbf{a}}(S_t - SS_t) \\ a_{0t} = 2S_t - SS_t \end{cases}$$

La prévision à l'horizon h est donnée par : $\hat{x}_{t+h} = a_{0t} + a_{1t} \times h$

Le Tableau 10 présente le calcul d'une prévision par lissage exponentiel double avec $\alpha = 0,3$.
Les Graphique 9 et

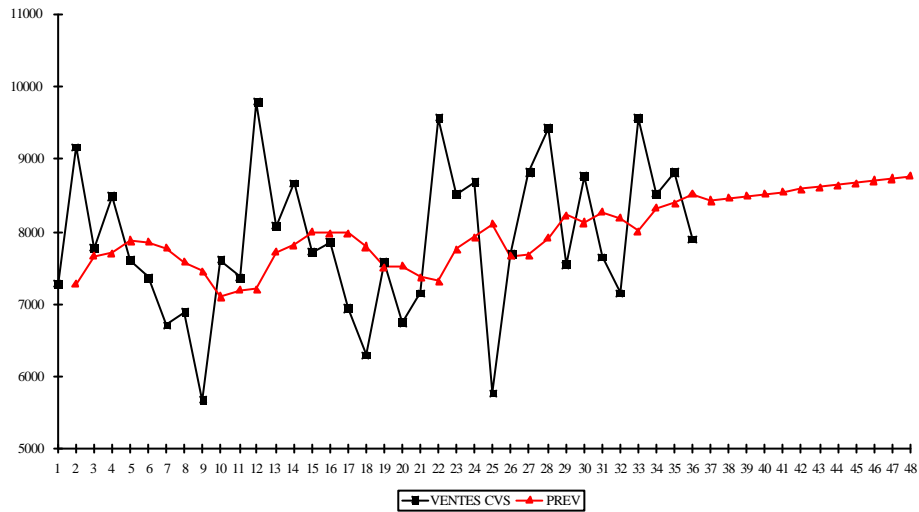
Graphique 10 illustrent une prévision par lissage exponentiel, sur les données, du Tableau 10 avec $\alpha = 0,1$ et $\alpha = 0,9$ (la prévision est fonction du poids accordé au passé récent et lointain).

Tableau 10 – Prévision par lissage exponentiel double avec $\alpha = 0,3$

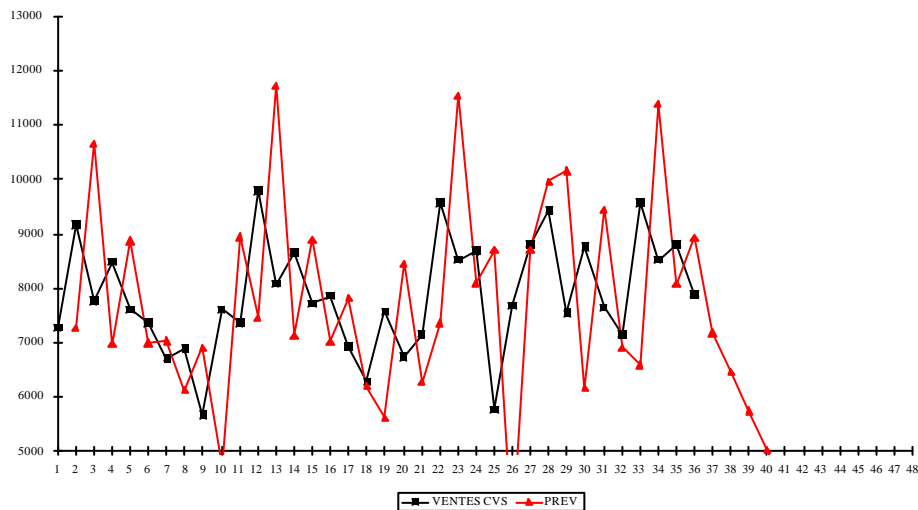
DATES	VENTES CVS	S	SS	a1t	a0t	PREV
Année 1-J	7281	7281,00	7281,00	0,00	7281,00	
F	9163	7845,60	7450,38	169,38	8240,82	7281,00
M	7769	7822,62	7562,05	111,67	8083,19	8410,20
A	8485	8021,33	7699,84	137,78	8342,83	8194,86
M	7606	7896,73	7758,91	59,07	8034,56	8480,62
J	7364	7736,91	7752,31	-6,60	7721,52	8093,63
J	6711	7429,14	7655,36	-96,95	7202,92	7714,92
A	6894	7268,60	7539,33	-116,03	6997,87	7105,97
S	5678	6791,42	7314,96	-224,37	6267,88	6881,84
O	7606	7035,79	7231,21	-83,75	6840,38	6043,51
N	7364	7134,26	7202,12	-29,09	7066,39	6756,63
D	9789	7930,68	7420,69	218,57	8440,67	7037,30
Année 2-J	8082	7976,07	7587,30	166,62	8364,85	8659,24
F	8663	8182,15	7765,76	178,45	8598,55	8531,46
M	7720	8043,51	7849,08	83,32	8237,93	8777,00
A	7859	7988,15	7890,80	41,72	8085,50	8321,25
M	6936	7672,51	7825,32	-65,49	7519,70	8127,23
J	6292	7258,36	7655,23	-170,09	6861,48	7454,21
J	7572	7352,45	7564,39	-90,83	7140,50	6691,40
A	6743	7169,61	7445,96	-118,43	6893,27	7049,67
S	7152	7164,33	7361,47	-84,49	6967,19	6774,83
O	9567	7885,13	7518,57	157,10	8251,69	6882,70
N	8522	8076,19	7685,86	167,29	8466,53	8408,79
D	8688	8259,73	7858,02	172,16	8661,45	8633,81
Année 3-J	5772	7513,41	7754,64	-103,38	7272,19	8833,61
F	7682	7563,99	7697,44	-57,19	7430,54	7168,81
M	8814	7938,99	7769,91	72,46	8108,08	7373,34
A	9432	8386,89	7955,00	185,10	8818,79	8180,54
M	7541	8133,13	8008,44	53,44	8257,81	9003,88
J	8765	8322,69	8102,72	94,27	8542,66	8311,25
J	7650	8120,88	8108,17	5,45	8133,60	8636,94
A	7152	7830,22	8024,78	-83,38	7635,65	8139,05
S	9567	8351,25	8122,72	97,94	8579,78	7552,27
O	8522	8402,48	8206,65	83,93	8598,30	8677,72
N	8814	8525,93	8302,43	95,79	8749,43	8682,23
D	7892	8335,75	8312,43	10,00	8359,08	8845,22
Année 4-J						8369,07
F						8379,07
M						8389,06
A						8399,06
M						8409,06
J						8419,05
J						8429,05
A						8439,04
S						8449,04
O						8459,04
N						8469,03
D						8479,03

Toutes les explications de calcul sont indiquées sur les cellules du fichier.

Graphique 9 – Prédiction par lissage exponentiel avec $\alpha = 0,1$



Graphique 10 – Prédiction par lissage exponentiel avec $\alpha = 0,9$



La lecture des deux graphiques précédents illustre parfaitement la technique du lissage exponentiel à plusieurs égards :

– d'une manière générale, nous observons que la prédiction réagit avec retard à une modification de la chronique ; c'est le phénomène de « course poursuite » entre réalisation et prédiction.

- La courbe tracée à partir du coefficient $\alpha = 0,8$ est très réactive à ces modifications brusques ; cependant, en prévision, elle amplifie fortement la tendance liée à la dernière observation.
 - Les deux prévisions calculées sont très différentes, la modification du coefficient α entraîne des résultats hétérogènes et instables.
- Ces graphiques présentent parfaitement les limites du lissage exponentiel.

4. Les modèles de Holt et Holt–Winters

4.1. Le modèle de Holt

La méthode que nous venons de présenter est celle de Brown. Nous pouvons aussi utiliser le lissage de Holt qui comprend deux paramètres : l'un pour la moyenne lissée a_{0t} et l'autre pour la pente a_{1t} .

Deux lissages distincts sont effectués :

- le lissage de la moyenne a_{0t} avec un coefficient de lissage α , $\alpha \in [0 ; 1]$,
- le lissage de la tendance a_{1t} avec un coefficient de lissage β , $\beta \in [0 ; 1]$.

(Dans le cas particulier où $\alpha = \beta$ le modèle de Holt se ramène au lissage exponentiel double de Brown).

Formulation

Lissage de la moyenne : $a_{0t} = \alpha x_t + (1 - \alpha) (a_{0t-1} + a_{1t-1})$

($a_{0t-1} + a_{1t-1}$ représente la nouvelle moyenne lissée en t)

Lissage de la tendance : $a_{1t} = \beta (a_{0t} - a_{0t-1}) + (1 - \beta) a_{1t-1}$

Prévision calculée en t à un horizon de h périodes : $\hat{x}_{t+h} = a_{0t} + h a_{1t}$

avec :

x_t = valeur observée de la série en t

Initialisation (pour $t = 1$)

- Initialisation de la moyenne lissée : $a_{01} = x_1$
- Initialisation de la tendance : $a_{11} = 0$

Les formules générales peuvent ensuite être utilisées.

Nous vous proposons à titre d'exercice de charger le fichier Web.

Un exemple complet de calcul est présenté, par la suite, à partir du modèle de Holt–Winters.

4.2. Le modèle avec tendance et saisonnalité (modèle de Holt–Winters)

Le modèle de Holt–Winters présente l'avantage d'intégrer une composante saisonnière et donc de réaliser le calcul de la prévision en un seul traitement. C'est ce modèle qui est employé le plus couramment dans les progiciels de prévision des ventes. Trois lissages¹⁵ distincts sont effectués :

- le lissage de la moyenne avec un coefficient de lissage α , avec $\alpha \in [0 ; 1]$,
- le lissage de la tendance avec un coefficient de lissage β , avec $\beta \in [0 ; 1]$,
- le lissage de la saisonnalité avec un coefficient de lissage γ , avec $\gamma \in [0 ; 1]$,

¹⁵ Si la série n'est pas saisonnière, on utilise alors le modèle de Holt à deux lissages : moyenne et tendance.

Formulation

Lissage de la moyenne : $a_{0t} = \alpha (x_t / S_{t-p}) + (1 - \alpha) (a_{0t-1} + a_{1t-1})$
(On utilise S_{t-p} car S_t n'est pas encore connue).

Lissage de la tendance : $a_{1t} = \beta (a_{0t} - a_{0t-1}) + (1 - \beta) a_{1t-1}$

Lissage de la saisonnalité : $S_t = \gamma (x_t / a_{0t}) + (1 - \gamma) S_{t-p}$

Prévision à un horizon de h périodes : $\hat{x}_{t+h} = (a_{0t} + h a_{1t}) S_{t-p+h}$ si $1 \leq h \leq p$
 $\hat{x}_{t+h} = (a_{0t} + h a_{1t}) S_{t-p+2h}$ si $p + 1 \leq h \leq 2p$

avec :

a_{0t} = moyenne lissée de la série en t

x_t = valeur observée de la série en t

S_t = coefficient saisonnier en t

p = périodicité des données ($p = 12$ en mensuel, $p = 4$ en trimestriel)

a_{1t} = tendance estimée en t .

Initialisation (pour la première année, $t = 1, p$)

– Initialisation de la saisonnalité

Les coefficients saisonniers pour la première année sont estimés par la valeur observée en t (x_t) divisée par la moyenne \bar{x} des p premières observations (celles de la première année).

$S_t = x_t / \bar{x}$ pour $t = 1, p$

– Initialisation de la moyenne lissée : $a_{0p} = \bar{x}$

– Initialisation de la tendance : $a_{1p} = 0$

Le Tableau 11 illustre un calcul de prévision à l'aide du modèle de Holt–Winters. Les données concernent les ventes d'un produit festif (vin de champagne).

Tableau 11 – Calcul d’une prévision par le modèle de Holt–Winters

avec $\alpha = 0,3$; $\beta = 0,1$; $\gamma = 0,2$

DATES	VENTES	a_{0t}	a_{1t}	S_t	\hat{x}_t
Année 1–	401,60			0,70	
F	395,70			0,69	
M	451,00			0,79	
A	427,60			0,75	
M	496,80			0,87	
J	467,70			0,82	
J	352,30			0,62	
A	182,10			0,32	
S	522,20			0,91	
O	687,20			1,20	
N	1080,3			1,89	
D	1391,6	571,3	0,0	2,44	
Année 2–	263,90	512,6	-5,9	0,67	
F	289,90	480,3	-8,5	0,67	350,93
M	337,00	458,3	-9,9	0,78	372,38
A	374,00	463,8	-8,3	0,76	335,61
M	292,70	419,8	-11,9	0,84	396,06
J	398,60	431,6	-9,5	0,84	333,94
J	421,70	500,6	-1,7	0,66	260,28
A	173,80	512,9	-0,3	0,32	159,04
S	522,10	530,2	1,5	0,93	468,51
O	642,40	532,4	1,6	1,20	639,48
N	984,20	529,9	1,1	1,88	1009,59
D	1307,6	532,8	1,3	2,44	1293,50
Année 3–	393,40	551,3	3,0	0,67	355,35
F	316,20	528,6	0,5	0,66	374,05
M	428,60	535,5	1,1	0,78	411,91
A	467,60	560,2	3,5	0,77	407,82
M	501,00	574,6	4,6	0,84	470,70
J	487,40	579,5	4,6	0,84	486,21
J	463,30	618,9	8,1	0,68	386,56
A	165,90	593,1	4,7	0,31	202,37
S	595,10	610,8	6,0	0,94	554,84
O	698,10	605,8	4,9	1,19	742,35
N	1012,10	588,6	2,7	1,85	1150,55
D	1380,00	583,6	1,9	2,42	1442,40
Année 4–					395,22
F					387,41
M					461,43
A					458,21
M					499,74
J					499,82
J					405,46
A					188,16
S					563,23
O					719,31
N					1119,40
D					1470,68

Exemples de calculs¹⁶.

Initialisation : $\bar{x} = 571,34$ (pour la première année).

La saisonnalité : $S_{avr-Année 1} = 427,60 / 571,34 = 0,75$

La moyenne : $a_{0déc-Année 1} = 571,34$

La tendance : $a_{1déc-Année 1} = 0$

¹⁶ Des légères différences peuvent apparaître entre les valeurs calculées et les valeurs du tableau du fait que le tableur utilise un nombre de décimales important.

Ici, en régime permanent (l'horizon h est choisi égal à 1) pour la ligne de septembre de l'année 2

:

$$a_{0,sep-Année 2} = 0,3 (522,1 / 0,91) + 0,7 (512,9 - 0,3) = 530,2$$

$$a_{1,sep-Année 2} = 0,1 (530,2 - 512,9) + 0,9 \times -0,3 = 1,5$$

$$S_{sep-Année 2} = 0,2 (522,1 / 530,2) + 0,8 \times 0,91 = 0,93$$

$$\hat{x}_{sep-Année 2} = (512,9 + (-0,3) \times 1) 0,91 = 468,51 \text{ (calculée en août 1992, avec } h = 1).$$

La prévision pour septembre de l'année 4 (horizon $h = 9$), calculée en décembre de l'année 3, est égale à: $\hat{x}_{sep-Année 4} = (583,6 + 1,9 \times 9) 0,94 = 563,23$

5. Comment choisir le coefficient de lissage ?

5.1. Principes généraux

Pour débiter le processus de lissage, il convient de choisir une valeur pour la constante α (par exemple $\alpha = 0,3$). Ce choix est très important car il conditionne la prévision future à travers le degré de pondération que l'on affecte au passé récent et au passé lointain.

Diverses procédures d'estimation de α ont été établies ; la plus classique consiste à retenir une valeur de α qui minimise l'écart entre la prévision et la réalisation sur la partie connue de la chronique (cf. 5.2.). Une autre approche consiste à élaborer des procédures de régulation et de contrôle qui permettent de modifier la constante du lissage. Ainsi, en cas de divergence systématique de la prévision, α s'ajuste automatiquement en vue de s'adapter à ce changement de structure (cf. 5.3).

5.2. Valeur de α qui minimise la somme des carrés des erreurs de prévision

C'est la technique de calcul la plus couramment employée, son principe est simple : pour un intervalle donné de valeurs de α ($\alpha_1 ; \alpha_2$) avec un « pas » assez fin (0,05 par exemple), les prévisions sont simulées et la somme des carrés des erreurs de prévision est alors calculée. Nous retenons la valeur de α qui rend minimum la somme des carrés des écarts. Cette technique peut être généralisée pour la détermination des trois coefficients (α, β, γ). L'utilisation du « SOLVEUR » sur les tableurs permet sans trop de difficultés de résoudre ce problème.

La démarche peut être illustrée de la manière suivante :

On cherche la valeur du coefficient de lissage qui minimise la somme des carrés des erreurs de

prévision passée : $\text{Min} \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (x_t - \hat{x}_t)^2$ en faisant varier la valeur de α tel que :

$$\alpha_1 = 0,1 ; \alpha_2 = 0,6 ; \text{pas} = 0,05$$

Soit $\alpha = 0,1 ; \alpha = 0,15 ; \alpha = 0,20 ; \dots ; \alpha = 0,6$

Le coefficient α retenu est celui qui correspond au minimum de $\sum_{t=1}^n (x_t - \hat{x}_t)^2$.

A titre d'illustration, recherche des coefficients de lissage optimaux dans un modèle de Holt-Winters cf. Fichier Web.

Cette technique de détermination du coefficient de lissage est optimale selon le critère des moindres carrés. Si nous sommes certains de trouver le meilleur coefficient sur l'ensemble de la période de l'estimation (c'est-à-dire sur le passé), nous ne sommes plus du tout certains d'être à l'optimum au moment où la prévision est calculée. La méthode suivante pallie cet inconvénient.

5.3. Procédure de régulation d'un coefficient de lissage

5.3.1 Principes généraux

La valeur optimale¹⁷ résulte d'un compromis entre l'inertie liée à l'intégration de données lointaines et la sensibilité aux valeurs récentes.

En cas d'erreur de prévision constatée, deux interprétations sont possibles :

- il s'agit d'un accident, le coefficient α doit alors diminuer afin de gommer l'effet de cette valeur anormale ;
- il s'agit d'une rupture de tendance durable, le coefficient α doit être augmenté afin d'intégrer plus rapidement cette rupture.

L'arbitrage est délicat entre ces deux hypothèses.

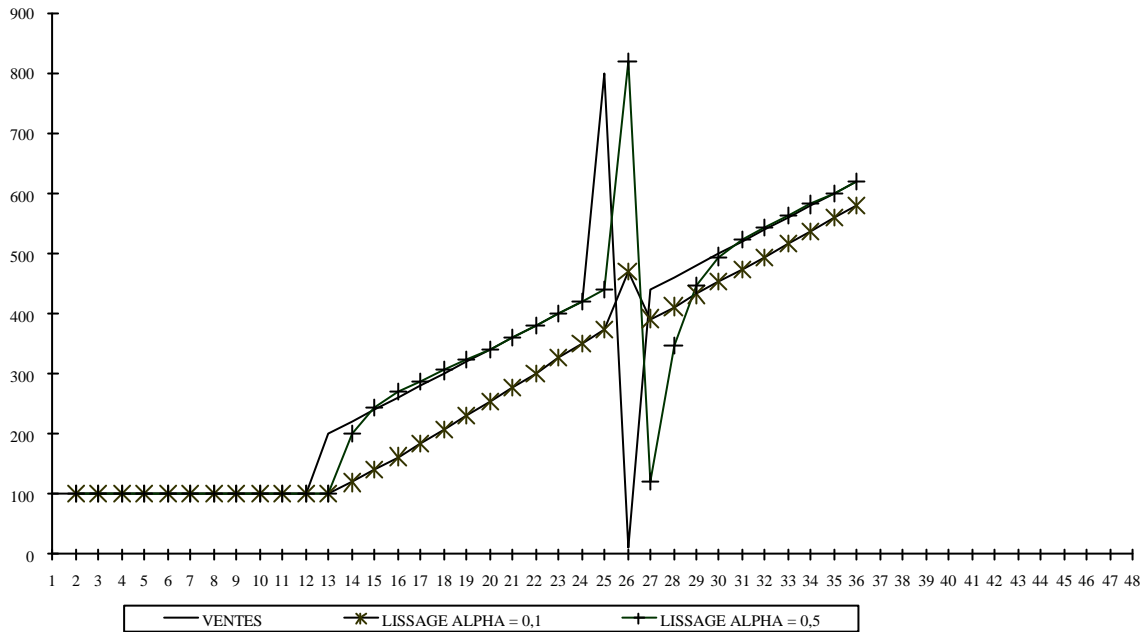
Prenons l'exemple d'une chronique de ventes qui connaît deux perturbations :

- une rupture de tendance à partir de la période 13,
- une valeur anormale pour les périodes 25 et 26.

Nous calculons une prévision (cf. Graphique 11) pour cette chronique à l'aide d'un lissage exponentiel double avec $\alpha = 0,1$ et $\alpha = 0,5$.

¹⁷ Si tant est qu'elle puisse exister, c'est-à-dire qu'un critère stable d'optimalité puisse être défini. A chaque nouvelle réalisation, la réestimation des paramètres tenant compte de l'erreur de prévision définit un nouveau coefficient de lissage. Il s'agit alors d'une « course poursuite » entre les prévisions et l'erreur de prévision, sans possibilité d'anticipation de la nouvelle structure de la chronique. On doit donc parler de valeur optimale instantanée.

Graphique 11 – Exemple de prévision avec $\alpha = 0,1$ et $\alpha = 0,5$



Si $\alpha = 0,1$: la réaction au changement de tendance et le réajustement sont longs. En revanche, ce coefficient permet d'effacer de façon très satisfaisante la perturbation,

Si $\alpha = 0,5$: le système s'ajuste bien et vite à la nouvelle moyenne mais sa sensibilité le fait réagir trop fort à la perturbation.

L'idéal est donc un coefficient susceptible d'évoluer au cours du temps ! La procédure de contrôle du coefficient α répond à cette question.

Reprenons le modèle du lissage exponentiel simple défini par l'équation :

$$\hat{x}_t = \alpha x_t + (1 - \alpha) \hat{x}_{t-1}$$

x_t = réalisation de la chronique (les ventes) x à la période t (un mois donné par exemple),

\hat{x}_t = prévision de la chronique x_t pour $t + 1$ calculée en t ,

α = coefficient de lissage, avec $\alpha \in [0 ; 1]$.

On est naturellement conduit à essayer de réguler de façon automatique et à chaque instant t , le paramètre α , afin de tenir compte, soit du passé immédiat, soit des valeurs rétrospectives plus anciennes.

On envisage un modèle du type :

$$\hat{x}_t = \alpha_t x_t + (1 - \alpha_t) \hat{x}_{t-1} \text{ dans lequel } \alpha_t \text{ évolue à chaque instant } t \text{ par un système de régulation.}$$

5.3.2 Les variables de contrôle

Les notations sont les suivantes :

x_1, x_2, \dots, x_t sont les valeurs successives de la chronique qu'il s'agit de prévoir aux différentes dates $1, 2, \dots, t$;

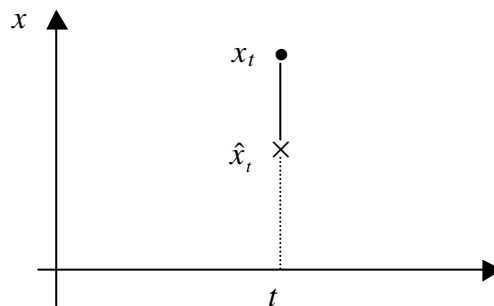
$\hat{x}_1, \hat{x}_2, \dots, \hat{x}_t$ sont les valeurs qui ont été prévues pour ces mêmes dates par la procédure de prévision.

Sur les différents graphiques, les croix représentent les réalisations x_t ; les points représentent les prévisions \hat{x}_t .

– *Indicateur d'écart instantané (EPS_t)*

L'écart entre prévision et réalisation (cf. Graphique 12) à l'instant t est la première variable qui vient à l'esprit pour juger de la performance du système de prévision : $EPS_t = x_t - \hat{x}_t$

Graphique 12 – Indicateur d'écart instantané



– *Indicateur de valeur cumulée des écarts (SUMEPS, somme des EPS)*

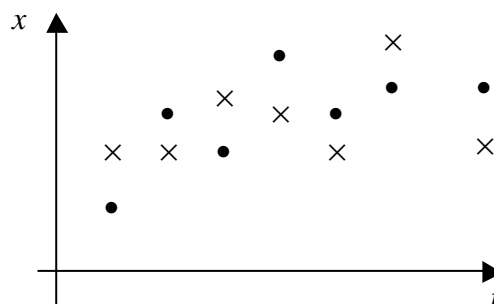
Soit T périodes de prévision successives (par exemple les 12 dernières prévisions) :

Il est défini par : $|SUMEPS_t| = \sum_{i=t-T+1}^t EPS_i$

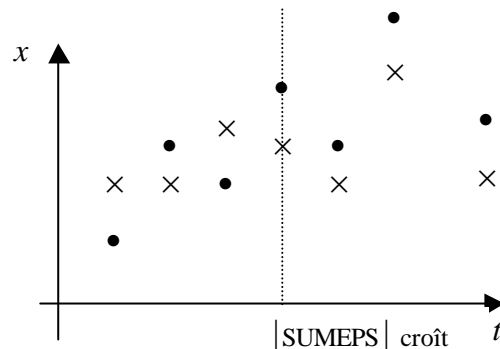
Si les observations sont – réparties – régulièrement autour des prévisions, la valeur de $SUMEPS$ est voisine à chaque instant de 0 (Graphique 13).

Si, en revanche, à partir d'un certain moment, $SUMEPS$ croît en valeur absolue, cela indique un changement de tendance ou de niveau (Graphique 14).

Graphique 13 – Erreur de prévision répartie



Graphique 14 – Prédiction biaisée



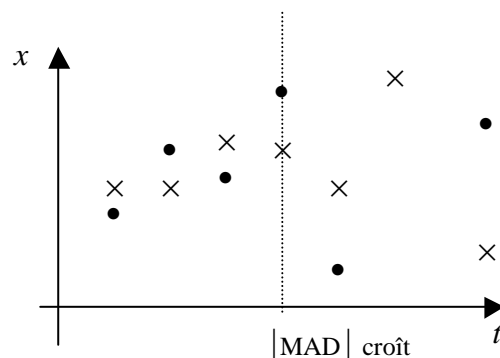
– Indicateur de la valeur absolue moyenne des écarts (*Mean Absolute Deviation*)

Il est défini à l'instant t par :
$$MAD_t = \frac{\sum_{i=t-T+1}^t |EPS_i|}{T}$$

Il s'agit d'une caractéristique de dispersion : MAD_t est l'analogue d'un écart type instantané des écarts de prévision.

Si les écarts EPS deviennent de plus en plus importants, la valeur de MAD s'accroît. Une erreur de prévision anormalement élevée entraîne une augmentation instantanée de la valeur MAD , cela correspond souvent à une observation aberrante (ou anormale) (cf. Graphique 15).

Graphique 15 – MAD croît avec t



Un écart instantané EPS_t ne peut se comprendre que par rapport aux fluctuations moyennes de la série (un écart instantané de 5 à 10 % peut paraître satisfaisant pour une série dont les oscillations sont violentes, mais anormal si la série est stable).

C'est pourquoi, il est souhaitable de comparer systématiquement EPS et $SUMEPS$ à l'indicateur d'écart moyen MAD , ce qui conduit à deux signaux de régulation NF et AWS .

– *Signal NF (Normal Forecast)*

Il est défini par : $|NF_t| = \frac{EPS_t}{MAD_t}$

Ce signal est l'analogie d'une variable centrée réduite (que l'on obtiendrait en remplaçant MAD_t , dans la formule précédente, par l'écart-type des écarts de prévision).

Cet indicateur est très sensible aux observations anormales (si $|EPS_t|$ est particulièrement élevé, $|NF_t|$ est aussi élevé).

– *Signal AWS (Alert Warning Signal)*

Il est défini par : $|AWS_t| = \frac{SUMEPS_t}{MAD_t}$

Dans le cas d'un changement de tendance, la somme algébrique des écarts de prévision est de même signe sur plusieurs périodes : la prévision ne s'ajuste pas convenablement au niveau des réalisations, le filtre $SUMEPS$ cumule ces écarts, $|AWS|$ augmente.

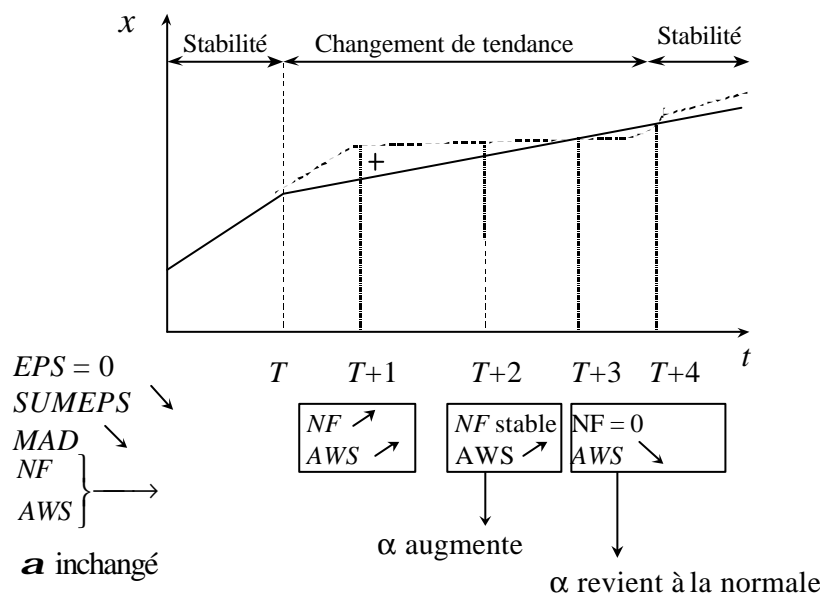
5.3.3 Application

– *Cas d'un changement de tendance* (Graphique 16)

Jusqu'à la période T , aucun indicateur n'augmente, α reste stable.

A partir de $T + 1$, la somme algébrique des erreurs est de même signe sur plusieurs périodes : la prévision est biaisée. Le filtre $SUMEPS$ va considérer ces écarts comme systématiques, NF augmente à la première période du changement de structure et peut par la suite, soit rester stable, soit maintenir son évolution ; AWS est le signal d'écart qui déclenchera l'augmentation de α , en considérant que la pondération accordée à la dernière information est trop faible.

Graphique 16 – Changement de tendance

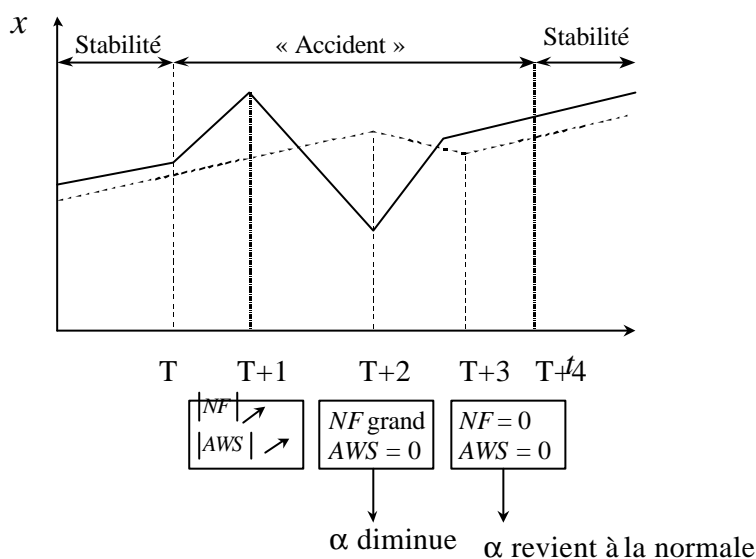


– Incident de type conjoncturel (Graphique 17)

Le phénomène a un caractère éphémère. Dans ce cas, les erreurs n'ont pas un caractère systématique mais peuvent être considérées comme anormales (filtre *MAD*). *AWS* augmente à la première période. Il peut demeurer stable, ou même s'annuler par la suite dans le cas d'écarts successifs importants de signes différents.

NF est le signal d'écart qui déclenche une diminution de α en considérant que la pondération accordée à la dernière information a un poids trop important puisqu'il s'agit d'un accident.

Graphique 17 – Accident conjoncturel



– La prévision a rejoint la nouvelle tendance où la valeur anormale est passée

Dans ce cas, α , qui a pu subir des variations importantes précédemment, revient au niveau de départ (α de base).

– Procédure opérationnelle

L'indicateur *NF* est donc spécifique d'une valeur anormale (en cas d'augmentation de *NF*, il convient de diminuer le coefficient de lissage) et l'indicateur *AWS* permet de détecter les ruptures de tendance (en cas d'augmentation de *AWS*, le coefficient de lissage doit être augmenté).

La question se pose alors de combiner d'une manière pertinente ces deux informations.

La constitution du tableau de décision (cf. Tableau 12) permet de répondre à ce problème. A partir de seuils expérimentaux¹⁸, nous définissons trois classes pour les valeurs de *NF* (en ligne) et *AWS* (en colonne) :

- *NF* est faible si sa valeur est inférieure à 1,5 \Rightarrow le code *NF* est alors 1,
- *NF* est moyen si sa valeur est comprise entre 1,5 et 3 \Rightarrow le code *NF* est 2,
- *NF* est fort si sa valeur est supérieure à 3 \Rightarrow le code *NF* est 3,
- *AWS* est faible si sa valeur est inférieure à 3 \Rightarrow le code *AWS* est 3,
- *AWS* est moyen si sa valeur est comprise entre 3 et 5 \Rightarrow le code *AWS* est 2,

¹⁸ C'est-à-dire issus d'expériences en entreprise.

– AWS est fort si sa valeur est supérieure à 5 \Rightarrow le code AWS est 1.

La stratégie à adopter – en ce qui concerne la valeur de α – se trouve à l'intersection d'une ligne et d'une colonne :

- si NF est faible et AWS fort, nous sommes confrontés à une rupture de tendance, le coefficient α doit fortement augmenter,
- si NF est fort et AWS moyen, nous avons une présomption d'une valeur anormale, le coefficient α doit légèrement diminuer,
- etc.

Tableau 12 – Table de décision en fonction des valeurs de NF et AWS

NF_t \ AWS_t	FAIBLE < 1,5 CODE = 1	MOYEN] 1,5 ; 3 [CODE = 2	FORT > 3 CODE = 3
FAIBLE < 3 CODE = 3	α moyen régime permanent	α faible présomption de valeur anormale	α très faible valeur anormale
MOYEN] 3 ; 5 [CODE = 2	α fort présomption de rupture de tendance	α moyen attente	α faible présomption de valeur anormale
FORT > 5 CODE = 1	α très fort rupture de tendance	α fort présomption de rupture de tendance	α moyen rupture de tendance ou valeur anormale ?

La valeur du coefficient α correspondant à la stratégie se trouve à l'intersection des valeurs de NF et AWS . D'une manière opérationnelle et afin de pouvoir automatiser le pilotage du coefficient de lissage, il convient de calculer la somme des deux codes NF et AWS :

$$\text{SOMME} = \text{CODE } NF + \text{CODE } AWS$$

Puis la valeur du coefficient se déduit directement de cette valeur selon l'équation empirique suivante :

$$\alpha_t = 0,7 - 0,1 \times \text{SOMME}$$

Le Tableau 13 présente en fonction des différentes valeurs de la somme des codes NF et AWS la valeur déduite de α .

Tableau 13 – Valeurs de α en fonction de la somme des codes *NF* et *AWS*

SOMME	α
6	0,1
5	0,2
4	0,3
3	0,4
2	0,5

Dans le cas d'un modèle de Holt ou Holt–Winters, nous sommes confrontés au pilotage de deux ou trois coefficients. La stratégie est alors la suivante :

- le coefficient de lissage de la moyenne α est piloté comme indiqué ci-dessus,
- le coefficient de lissage de la tendance β , sur lequel on désire être légèrement moins réactif que celui de la moyenne, se déduit de la valeur de α par l'équation $\beta_t = \alpha_t - 0,05$,
- enfin, le coefficient γ de la saisonnalité, dans le cas du modèle de Holt–Winters, reste statique et égal à 0,2 car les ruptures de tendance ou les valeurs anormales n'affectent pas la composante saisonnière.

A titre d'illustration : Fichier Web (pilotage d'un coefficient dans un modèle de lissage double).

L'avantage des méthodes fondées sur le lissage exponentiel réside dans leur grande facilité de mise en œuvre et leur simplicité relative de calcul. En revanche, nous pouvons leur reprocher une rusticité qui s'avère suffisante en cas de chronique assez peu fluctuante, mais devient vite gênante si la série est fortement chahutée. De surcroît, par son inspiration même, l'environnement extérieur n'est pas pris en compte et n'a donc aucun effet. Cependant, dans certains cas il est possible d'intégrer, de manière rudimentaire, des facteurs explicatifs qui viennent alors *a priori* améliorer la prévision obtenue (voir chapitre 7).

En conclusion de ce chapitre, nous pouvons synthétiser les avantages et les inconvénients de cette technique de prévision à l'aide du Tableau 14.

Tableau 14 – Points forts et points faibles du lissage exponentiel

<i>Avantages</i>	<i>Inconvénients</i>
Simplicité des calculs pour l'élaboration des prévisions en régime permanent	Méthode purement extrapolative, qui ignore les événements exogènes
Système adaptatif avec calculs itératifs, facilement réalisable sur tableur	Choix délicat de la constante de lissage
Rapidité de calcul	Système réagissant avec retard
Maîtrise du fonctionnement par l'utilisateur	

Chapitre 4

PRINCIPALES METHODES (2) :

BOX-JENKINS ET MODELE CAUSAL

Dans ce chapitre nous étudions deux méthodes de prévision :

- La méthodologie de Box et Jenkins est une technique d'extrapolation par référence à une loi de reproduction du phénomène étudié. L'information permettant de mettre en évidence le processus est contenue dans la série chronologique elle-même, sans apport externe, d'où le nom de prévision endogène ;
- les modèles causals ou modèles explicatifs cherchent à expliquer les fluctuations de la série des ventes par référence à des facteurs explicatifs externes représentatifs soit de la politique marketing de la firme, soit de la demande.

1. La corrélation et le corrélogramme

Deux concepts doivent tout d'abord être explicités : la corrélation et le corrélogramme.

1.1. La corrélation

Lorsque deux phénomènes ont une évolution commune, ils sont dits « corrélés ». La corrélation simple mesure le degré de liaison existant entre ces deux phénomènes représentés par des variables x et y .

Nous pouvons distinguer la corrélation linéaire, lorsque tous les points du couple de valeurs (x , y) des deux variables semblent alignés sur une droite, de la corrélation non linéaire lorsque le couple de valeurs se trouve sur une même courbe d'allure quelconque.

Deux variables peuvent être :

- en corrélation positive, on constate une augmentation (ou diminution, ou constance) simultanée des valeurs des deux variables ;
- en corrélation négative, lorsque les valeurs de l'une augmentent, les valeurs de l'autre diminuent ;
- non corrélées, lorsqu'il n'y a aucune relation entre les variations des valeurs de l'une des variables et les valeurs de l'autre.

1.1.1 Mesure du coefficient de corrélation linéaire

La représentation graphique ne donne qu'une « impression » de la corrélation entre deux variables sans donner une idée précise de l'intensité de la liaison ; c'est pourquoi nous calculons une statistique appelée coefficient de corrélation linéaire simple noté r_{xy} . Il est égal à :

$$r_{xy} = \frac{Cov(x, y)}{s_x \times s_y} = \frac{\sum_{t=1}^{t=n} (x - \bar{x})(y - \bar{y})}{\sqrt{\left(\sum_{t=1}^{t=n} (x - \bar{x})^2\right)} \sqrt{\left(\sum_{t=1}^{t=n} (y - \bar{y})^2\right)}}$$

$$= \frac{n \sum_{t=1}^{t=n} x_t y_t - \sum_{t=1}^{t=n} x_t \sum_{t=1}^{t=n} y_t}{\sqrt{n \sum_{t=1}^{t=n} x_t^2 - \left(\sum_{t=1}^{t=n} x_t\right)^2} \sqrt{n \sum_{t=1}^{t=n} y_t^2 - \left(\sum_{t=1}^{t=n} y_t\right)^2}}$$

avec :

x_t et y_t = valeurs des deux historiques à l'instant t ,

$Cov(x,y)$ = covariance entre x et y ,

S_x et S_y = écart-type de x et écart-type de y ,

\bar{x} et \bar{y} = moyenne de x et moyenne de y ,

n = nombre d'observations.

On peut démontrer que, par construction, ce coefficient reste compris entre -1 et 1 :

- proche de 1, les variables sont corrélées positivement,
- proche de -1, les variables sont corrélées négativement,
- proche de 0, les variables ne sont pas corrélées.

Dans la pratique, ce coefficient est rarement très proche de l'une de ces trois bornes et il est donc difficile de proposer une interprétation fiable à la simple lecture de ce coefficient. La théorie des tests statistiques nous permet de lever cette indétermination. Nous calculons un ratio (appelé le t de Student empirique) :

$$t_{cal} = \frac{r_{x,y}}{\sqrt{\frac{1 - r_{x,y}^2}{n-2}}}$$

Si $t_{cal} > t_{lu}$ lu dans une table de Student¹⁹ à $n - 2$ degrés de liberté (si le nombre d'observations est supérieur à 30, on peut prendre par approximation $t_{lu} = 2$), le coefficient de corrélation est significativement différent de 0 ; dans le cas contraire, l'hypothèse d'un coefficient de corrélation nul est acceptée.

Que veut dire significativement différent de 0 ?

On cherche à déterminer le risque de se tromper en affirmant que le coefficient (vrai et inconnu) est différent de 0. Ce risque admissible est en général de 5 % (on accepte de se tromper dans 5% des cas) ; mais plus le seuil de risque fixé est faible, plus forte est notre certitude quant à l'existence de la corrélation entre les deux variables.

La liaison est en général considérée comme significative si on a un risque de se tromper inférieur à 5% en affirmant que le coefficient de corrélation est bien différent de 0.

Le Tableau 15 présente un exemple de calcul d'un coefficient de corrélation entre les ventes trimestrielles d'un produit (x_t) et les dépenses publicitaires (y_t) trimestrielles.

¹⁹ Les tables statistiques sont en fin d'ouvrage.

Tableau 15 – Exemple de calcul d'un coefficient de corrélation

Trimestre	x	y	x ²	y ²	x y
1	16	20	256	400	320
2	18	24	324	576	432
3	23	28	529	784	644
4	24	22	576	484	528
5	26	32	676	1024	832
6	28	32	784	1024	896
7	29	28	841	784	812
8	31	36	961	1296	1116
9	32	41	1024	1681	1312
10	34	41	1156	1681	1394
Somme	261	304	7127	9734	8286

$$r_{x,y} = \frac{(10)(8286) - (261)(304)}{\sqrt{(10)(7127) - 261^2} \sqrt{(10)(9734) - 304^2}} = 0,89$$

$$t_{cal} = \frac{0,89}{\sqrt{\frac{1 - 0,89^2}{10 - 2}}} = 5,49 > t_{lu} = 2,306 \text{ (8 degrés de liberté et } \alpha = 0,05\text{)}.$$

La corrélation est donc significative.

1.1.2 Limites de la notion de corrélation

a) Corrélation n'est pas causalité

Le fait d'avoir un coefficient de corrélation élevé entre deux variables ne signifie pas qu'il existe un lien de causalité. En d'autres termes, une covariance significativement différente de 0 n'implique pas une liaison d'ordre économique, physique ou autres. Nous appelons corrélation fortuite ce type de corrélation que rien ne peut expliquer.

L'exemple le plus fameux concerne la forte corrélation existant entre le nombre de taches solaires observées et le taux de criminalité aux Etats-Unis. Cela ne signifie pas qu'il existe une relation entre les deux variables, mais qu'un troisième phénomène, l'évolution de long terme (la tendance) ici, explique conjointement les deux phénomènes.

b) La relation testée est linéaire

L'application de la formule de calcul ne permet de déterminer que des corrélations linéaires entre variables. Un coefficient de corrélation nul indique que la covariance entre la variable x et la variable y est égale à 0. C'est ainsi que deux variables en totale dépendance peuvent avoir un coefficient de corrélation nul, comme l'illustre l'exemple suivant : l'équation d'un cercle nous est donnée par $x^2 + y^2 = R^2$, les variables x et y sont bien liées entre elles par une relation et pourtant leur covariance est nulle et donc leur coefficient de corrélation est égal à 0.

Pour pallier cette limite, il convient éventuellement de transformer les variables, préalablement au calcul du coefficient de corrélation, afin de linéariser leur relation, par exemple, au moyen ici d'une transformation de type logarithmique.

1.2. Le corrélogramme et la fonction d'autocorrélation

L'autocorrélation est un concept lié à celui de corrélation : il s'agit non pas d'un calcul entre deux chroniques différentes mais entre la série et elle-même à différents décalages dans le temps.

On appelle coefficient d'autocorrélation d'ordre 1 le coefficient de corrélation linéaire calculé entre la série chronologique et cette même série décalée d'une période de temps. Les coefficients d'autocorrélation sont calculés pour des ordres allant de 0 à K , K étant le décalage maximum

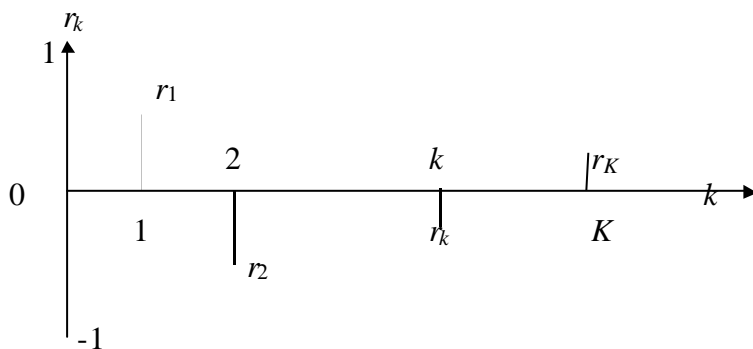
admissible (en général $\frac{n}{6} \leq K \leq \frac{n}{3}$ ou $K = \frac{n}{5}$ si $n \geq 150$) pour que le coefficient d'autocorrélation ait un sens (cf. Tableau 16).

Tableau 16 – Exemple de calcul d'une autocorrélation

	Retards	Autocorrélation
1 2 3 ... t ... n	0	$r_0 = 1$
1 2 3 ... $t-1$... $n-1$	1	r_1
1 2 3 $t-2$... $n-2$	2	r_2
...
$t-k$... $n-k$	k	r_k
...
$t-K$... $n-K$	K	r_K

La représentation graphique de la Fonction d'AutoCorrélation (notée FAC) est appelée le corrélogramme comme l'illustre le Graphique 18.

Graphique 18 – Exemple de corrélogramme



Le coefficient d'autocorrélation d'ordre k est donné par :

$$r_k = \frac{\sum_{t=k+1}^n (x_t - \bar{x}_1)(x_{t-k} - \bar{x}_2)}{\sqrt{\sum_{t=k+1}^n (x_t - \bar{x}_1)^2 \sum_{t=k+1}^n (x_{t-k} - \bar{x}_2)^2}} = \frac{\sum_{t=k+1}^n x_t x_{t-k} - (n-k) \bar{x}_1 \bar{x}_2}{\sqrt{\left(\sum_{t=k+1}^n x_t^2 - (n-k) \bar{x}_1^2\right) \left(\sum_{t=k+1}^n x_{t-k}^2 - (n-k) \bar{x}_2^2\right)}}$$

$$\bar{x}_1 = \frac{1}{n-k} \sum_{t=k+1}^n x_t \quad \bar{x}_2 = \frac{1}{n-k} \sum_{t=k+1}^n x_{t-k}$$

Le test de signification sur le coefficient r_k permet de sélectionner les coefficients d'autocorrélation significativement différents de 0 ; il s'effectue comme pour un coefficient de

corrélation linéaire simple. Soit le calcul de la quantité $t_c = \frac{|r_k|}{\sqrt{1-r_k^2}} \sqrt{n-2}$. Si $t_c > t_{n-2}^{\alpha/2}$, le

coefficient d'autocorrélation est significativement différent de 0 ($t_{n-2}^{\alpha/2}$ valeur de la loi de Student au seuil α à $n-2$ degrés de liberté).

Le calcul de r_k permet de déceler des liaisons internes à la série. Par exemple, une série saisonnière, avant désaisonnalisation, montre une valeur élevée pour r_{12} . A travers la connaissance des valeurs de la fonction d'autocorrélation, le processus de reproduction interne à la série peut être décelé. Pour toutes les méthodes endogènes, il s'agit donc d'une étape nécessaire, préalable à la modélisation.

Deux propriétés doivent être signalées :

$r_0 = 1$ (corrélation parfaite de la série avec elle-même à décalage zéro),

$r_{-k} = r_k$ (les valeurs du corrélogramme sont identiques pour deux décalages temporels symétriques par rapport à la période de référence).

Le corrélogramme est la représentation graphique de cette fonction d'autocorrélation.

Il est souhaitable de présenter sur ce graphe l'intervalle de confiance autour de zéro ($\pm 2/\sqrt{n}$), permettant d'éliminer visuellement tous les termes non significativement différents de zéro.

Ainsi, si tous les r_k se trouvent à l'intérieur de cette bande de confiance, la chronique dont le corrélogramme a été calculé est purement aléatoire ; c'est-à-dire qu'il s'avère impossible de trouver une loi statistique de reproduction du phénomène. Nous allons détailler les calculs, à partir des données de ventes d'une série trimestrielle (cf. Tableau 17), pour un coefficient, par exemple r_2 , c'est-à-dire la corrélation entre x_t et x_{t-2} .

Tableau 17 – Exemple de calcul pour un coefficient d'autocorrélation d'ordre 2

t	x_t	x_{t-2}	$(x_t - \bar{x}_1)(x_{t-2} - \bar{x}_2)$	$(x_t - \bar{x}_1)^2$	$(x_{t-2} - \bar{x}_2)^2$
1	1057	1248	195360,56	431162,99	88518,15
2	3159	1392	-222259,17	2090540,47	23629,84
3	891	1057	402349,24	676391,15	239336,23
4	1065	3159	-1045292,60	419813,28	2602672,82
5	1118	891	390149,56	354775,05	429051,24
6	2934	1065	-586700,47	1490767,59	230899,47
7	1138	1118	246239,35	330659,57	183372,34
8	1456	2934	-357132,95	66167,26	1927598,84
9	1224	1138	199297,67	239052,53	166154,12
10	3090	1456	-123679,42	1896046,27	8067,63
Somme	17132,30	15458,20	-901668,23	7995376,17	5899300,68

$$\bar{x}_1 = 1713,2 \text{ et } \bar{x}_2 = 1545,8$$

$$r_2 = \frac{\sum_{t=k+1}^n (x_t - \bar{x}_1)(x_{t-2} - \bar{x}_2)}{\sqrt{\sum_{t=k+1}^n (x_t - \bar{x}_1)^2} \sqrt{\sum_{t=k+1}^n (x_{t-2} - \bar{x}_2)^2}} = \frac{-901668,23}{\sqrt{5899300,68} \sqrt{7995376,17}} = -0,13$$

Le t calculé est égal pour $n = 10$ et $r_k = -0,13$: $t_c = \frac{|r_k|}{\sqrt{1-r_k^2}} \sqrt{n-2} = 0,38 < t_{10-2}^{0,05/2} = 2,306$. Le

coefficient n'est pas significativement différent de 0.

Le Tableau 18 indique l'ensemble des valeurs de la fonction d'autocorrélation que l'on compare à la valeur lue dans la table de Student pour un seuil de 5% et à $n - 2$ degrés de liberté. Seul le coefficient d'autocorrélation d'ordre 4 est significativement différent de 0 ; la périodicité des données étant trimestrielle, ce « pic » est donc attribué à la saisonnalité des données.

Tableau 18 – Calcul d'une fonction d'autocorrélation

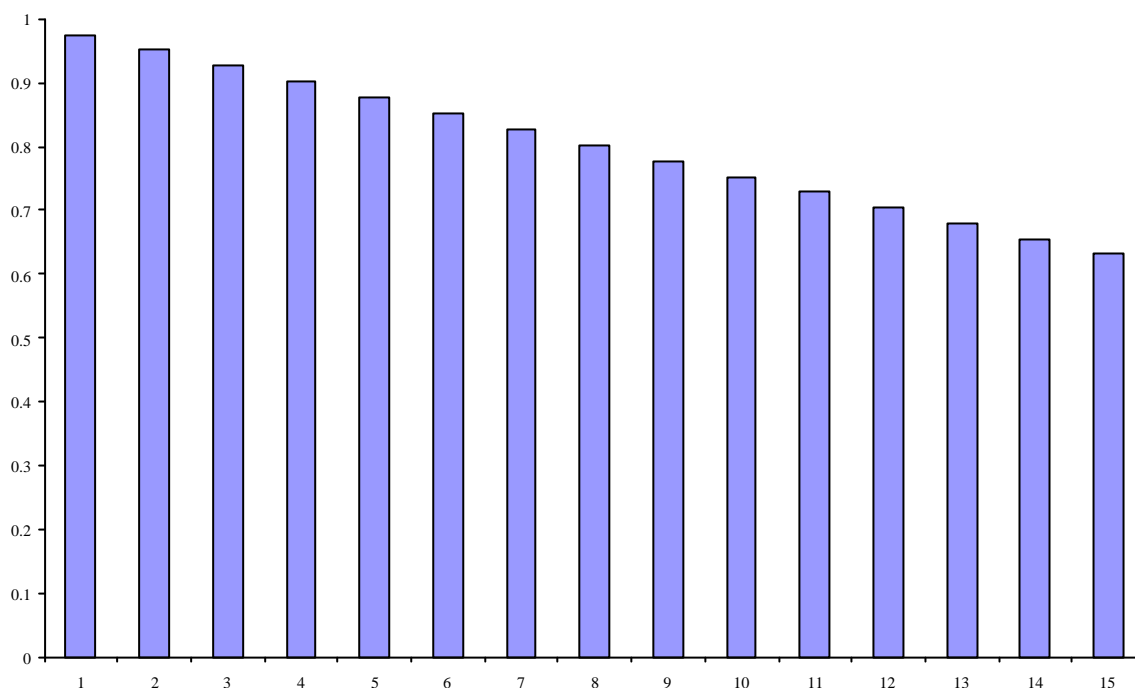
Retards	r_k	n	t_c	$d d l$	t_{lu} à 0,05
0	1	12	-	-	-
1	-0,395	11	1,29	9	2,262
2	-0,132	10	0,38	8	2,306
3	-0,392	9	1,13	7	2,365
4	0,952	8	7,62	6	2,447

Dans une approche endogène, la première étape consiste toujours à décomposer une chronique en trois composantes fondamentales : tendance, saisonnalité, résidu, puis à les extrapoler indépendamment les unes des autres et enfin à les agréger en vue d'obtenir une prévision en termes directement exploitables.

L'extrapolation de la tendance et de la saisonnalité ne pose pas de problème particulier ; en ce qui concerne l'extrapolation du facteur résiduel, de nombreuses méthodes existent, toutes fondées sur l'étude préalable du corrélogramme.

Si au lieu de traiter le facteur résiduel, on s'intéresse à la série désaisonnalisée (tendance et résidus), l'allure du corrélogramme est alors typique d'une chronique non stationnaire (la moyenne de la série est instable au cours du temps, car la tendance est incluse dans la série, cf. Graphique 19.) et nous ne pourrions obtenir aucune information exploitable. C'est pourquoi il est impératif d'effectuer les calculs de la fonction d'autocorrélation sur la série hors tendance et saisonnalité.

Graphique 19 – Corrélogramme d'une série non stationnaire (« *trendée* »)



Comme nous l'avons évoqué plus haut, il est possible qu'aucun terme du corrélogramme ne soit significativement différent de zéro. Ceci laisse présager un processus purement aléatoire et donc, par définition, imprévisible. La prévision ne peut alors être calculée qu'à l'aide d'une extrapolation de la tendance ou d'un lissage exponentiel. Dans le cas contraire, Box et Jenkins proposent une méthodologie permettant de déterminer le processus de reproduction sous-jacent à la série de vente.

1.3. Introduction à la méthodologie de Box et Jenkins

Dans le cas où les premiers termes du corrélogramme diffèrent de zéro, il convient de modéliser le terme résiduel, c'est-à-dire de trouver la loi de reproduction du phénomène.

Plusieurs modèles endogènes existent, plus ou moins complexes²⁰. Seule la méthode de Box-Jenkins, qui a fait l'objet d'importantes applications dans le domaine de la prévision, est présentée ici.

Box et Jenkins²¹ ont développé une véritable méthodologie de recherche systématique d'un modèle adéquat en fonction de l'étude des corrélogrammes empiriques, ils se réfèrent à deux types de modèle : des processus moyennes mobiles, des modèles autorégressifs, ou à une combinaison des deux.

1.3.1 Typologie des modèles

Les modèles ARMA permettent de représenter la plupart des processus aléatoires stationnaires (Wold, 1954) ; ainsi ils sont censés recouvrir une gamme très large d'évolutions possibles de séries chronologiques. Nous pouvons distinguer deux types de processus : les processus autorégressifs (AR) et les processus de moyenne mobile (MA). Chacun de ces modèles est caractérisé par sa fonction d'autocorrélation simple (FAC) et sa fonction d'autocorrélation partielle (FAP).

²⁰ Dans le cadre de cet ouvrage, nous nous limitons à des développements succincts, le lecteur intéressé par la méthode de Box et Jenkins peut se référer à Bourbonnais et Terraza (1998).

²¹ Box et Jenkins (1976).

La Fonction Autocorrélation Partielle (notée FAP) est une notion plus complexe que la FAC. Il s'agit de la corrélation entre x_t et x_{t-k} , l'influence des variables x_{t-k-i} pour ($i < k$) ayant été retirée.

– *Les modèles AR*

La partie autorégressive d'un processus, notée AR, est constituée par une combinaison linéaire finie des valeurs passées du processus. Le processus AR(p) est donc défini à partir de la formule générale : $x_t = f_1 x_{t-1} + \dots + f_p x_{t-p} + a_t$ avec a_t un processus de bruit blanc gaussien ; par exemple :

$$\text{AR}(1) : x_t = f_1 x_{t-1} + a_t ; \text{AR}(2) : x_t = f_1 x_{t-1} + f_2 x_{t-2} + a_t ; \text{etc.}$$

Un processus AR(p) a les caractéristiques suivantes :

- la FAC est une exponentielle et/ou une sinusoïde amortie,
- seuls les p premiers termes de la FAP sont significativement différents de 0.

– *Les modèles MA*

La partie moyenne mobile, notée MA, est constituée d'une combinaison linéaire finie en t des valeurs passées d'un bruit blanc. Un processus MA(q) est défini à partir de la formule :

$$x_t = a_t - q_1 a_{t-1} - \dots - q_q a_{t-q} . \text{ avec } a_t \rightarrow \text{processus de bruit blanc gaussien ; par exemple :}$$

$$\text{MA}(1) : x_t = a_t - q_1 a_{t-1} ; \text{MA}(2) : x_t = a_t - q_1 a_{t-1} - q_2 a_{t-2}$$

Un processus MA(q) a les caractéristiques suivantes :

- seuls les q premiers termes de la FAC sont significativement différents de 0,
- la FAP est une exponentielle et/ou une sinusoïde amortie.

Nous pouvons constater la symétrie parfaite des comportements de la FAC d'un AR(p) et de la FAP d'un MA(q) d'une part, et de la FAP d'un AR(p) et de la FAC d'un MA(q) d'autre part. En effet, nous pouvons démontrer l'équivalence suivante : AR(1) = MA(∞) et MA(1) = AR(∞).

– *Les modèles ARMA*

Les processus ARMA sont des mélanges de processus AR et MA ; ils sont donc définis par le modèle suivant :

$$\text{ARMA}(p, q) : x_t = f_1 x_{t-1} + \dots + f_p x_{t-p} + a_t - q_1 a_{t-1} - \dots - q_q a_{t-q} . \text{ Par exemple :}$$

$$\text{ARMA}(1, 1) : x_t = f_1 x_{t-1} + a_t - q_1 a_{t-1}$$

$$\text{ARMA}(2, 1) : x_t = f_1 x_{t-1} + f_2 x_{t-2} + a_t - q_1 a_{t-1} ; \text{etc.}$$

On remarque que : ARMA(0, q) \equiv MA(q) et ARMA(p , 0) \equiv AR(p)

Le Tableau 19 synthétise les caractéristiques, en termes de corrélogrammes, des processus AR, MA et ARMA.

Tableau 19 – Résumé des propriétés des fonctions d'autocorrélation simple et partielle

<i>Processus</i>	<i>FAC</i>	<i>FAP</i>
AR(1)	Décroissance exponentielle ($\alpha > 0$) ou sinusoidal amortie ($\alpha < 0$)	Pic significatif pour le premier retard: Positif si $\alpha > 0$ et négatif si $\alpha < 0$, les autres coefficients nuls pour des retards > 1
AR(2)	Décroissance exponentielle ou sinusoidal selon les signes de α et β	Pics significatifs pour le premier et second retards, les autres coefficients sont nuls pour des retards > 2
AR(p)	Décroissance exponentielle et/ou sinusoidal	Pics significatifs pour les p premiers retards, les autres coefficients sont nuls pour des retards $> p$
MA(1)	Pic significatif pour le premier retard : positif si $\theta < 0$ et négatif si $\theta > 0$. Les autres coefficients sont nuls pour des retards > 1	Décroissance exponentielle ($\theta > 0$) ou sinusoidal amortie ($\theta < 0$)
MA(2)	Pics significatifs pour le premier et second retards. Les autres coefficients sont nuls pour des retards > 2	Décroissance exponentielle ou sinusoidal selon les signes de θ et ϕ
MA(q)	Pics significatifs pour les q premiers retards. Les autres coefficients nuls pour des retards $> q$	Décroissance exponentielle et/ou sinusoidal
ARMA(1, 1)	Décroissance géométrique à partir du premier retard, le signe est déterminé par $\alpha - \theta$	Décroissance exponentielle ($\theta > 0$) ou sinusoidal amortie ($\theta < 0$)
ARMA(p, q)	Décroissance exponentielle ou sinusoidal amortie tronquée après ($q - p$) retards	Décroissance exponentielle ou sinusoidal amortie tronquée après $p - q$ retards

– *Conditions d'utilisation*

Les modèles AR, MA, ARMA ne sont représentatifs que de chroniques :

- stationnaires en moyenne (c'est-à-dire hors tendance),
- corrigées des variations saisonnières.

1.3.2 Le problème de la stationnarité

Nous ne pouvons interpréter valablement les corrélogrammes que sur des séries stationnaires (sans tendance). Les tests de Dickey-Fuller (1979) et Dickey-Fuller Augmenté (1981) permettent de déterminer si la série est stationnaire et, dans le cas d'une non stationnarité, de quel type il s'agit : TS (« *Trend Stationnary* ») qui représente une non stationnarité de type déterministe ou DS (« *Differency Stationnary* ») pour les processus non stationnaires stochastiques. Si la série étudiée est de type TS, il convient de la stationnariser par régression sur le temps ; le résidu d'estimation est alors étudié selon la méthodologie de Box-Jenkins. Ceci permet de déterminer les ordres p et q des parties AR et MA du résidu. Le modèle est toujours dans ce cas un ARMA(p, q).

Si la série étudiée est de type DS, il convient de la stationnariser par passage aux différences selon l'ordre d'intégration $I = d$ (d est le nombre de fois qu'il faut différencier la série pour la rendre stationnaire). La série différenciée est alors étudiée selon la méthodologie de Box-Jenkins qui permet de déterminer les ordres p et q des parties AR et MA. On note ce type de modèle ARIMA(p, d, q).

1.3.3 Recherche de la représentation adéquate

L'objectif est de rechercher parmi tous les modèles (AR, MA, ARMA) celui qui semble le mieux adapté, puis de spécifier le (ou les) degré(s) du processus retenu. Plusieurs étapes sont nécessaires.

– *L'identification*

La phase d'identification est la plus importante et la plus difficile : elle consiste à déterminer le modèle adéquat, c'est-à-dire les valeurs des paramètres p , d , q du modèle ARIMA. Elle est fondée sur l'étude des corrélogrammes simple et partiel.

Dans la méthodologie de Box et Jenkins, le problème de l'identification du modèle est donc crucial.

– *Estimation et validation*

L'estimation des paramètres du modèle est fondée sur la maximisation d'une fonction de vraisemblance. La validation de la représentation porte sur :

– Les coefficients du modèle qui doivent être significativement différents de 0 (le test de Student s'applique de manière classique). Si un coefficient n'est pas significativement différent de 0, il convient d'envisager une nouvelle spécification éliminant l'ordre du modèle AR ou MA non valide.

– L'analyse des résidus (écart entre la série observée et la série prévue) permet de vérifier qu'ils sont :

- de moyenne nulle ; dans le cas contraire, il convient d'ajouter une constante au modèle ;
- représentatif d'un bruit blanc. Si le résidu n'est pas un bruit blanc, cela signifie que la spécification du modèle est incomplète et qu'il manque au moins un ordre à un processus.

La phase de validation du modèle est très importante et nécessite le plus souvent un retour à la phase d'identification.

– *Prévision*

Lorsque le modèle est validé, la prévision peut alors être calculée à un horizon h de quelques périodes, limité par la variance de l'erreur de prévision qui augmente rapidement avec l'horizon.

A l'étape de la transformation, plusieurs techniques ont pu être employées afin de stationnariser le processus générateur pour qu'il soit identifiable dans la classe des processus ARMA. Il est nécessaire, lors de l'étape de prévision, de prendre en compte la ou les transformations retenues (« recoloration de la prévision »).

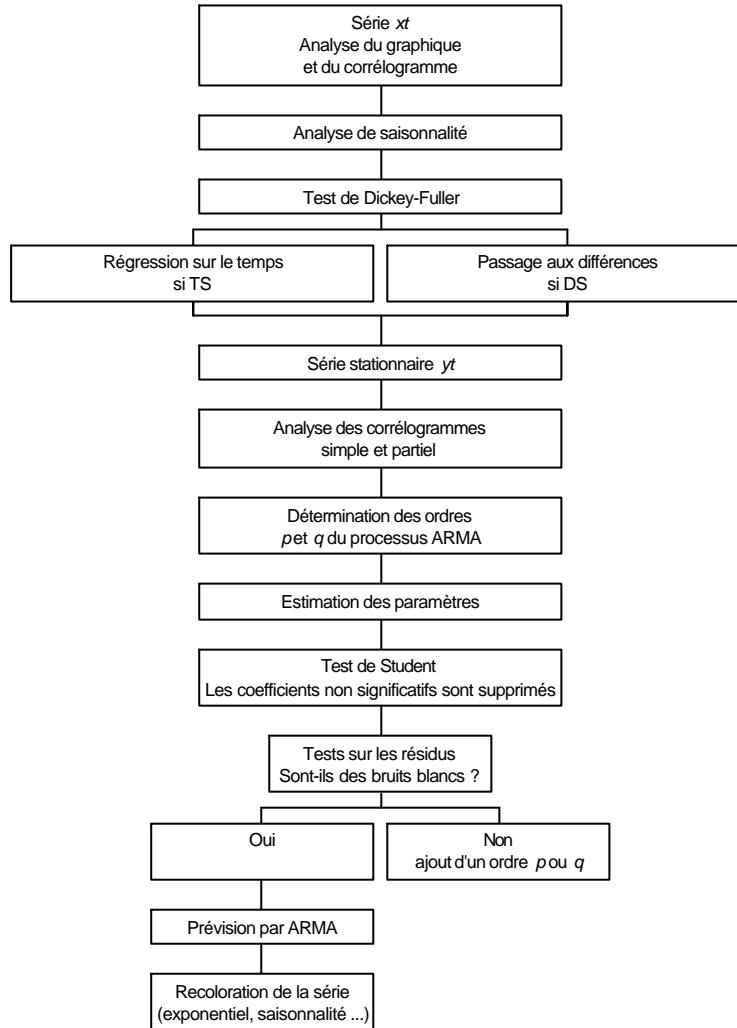
– Si on a utilisé la régression afin de retirer une ou plusieurs composantes déterministes estimées par les Moindres Carrés Ordinaires, ces dernières sont extrapolées jusqu'à l'horizon prévisionnel choisi puis combinées aux valeurs prévisionnelles de l'ARMA.

– Si la série a été préalablement désaisonnalisée, il convient de re-saisonnaliser la série des ventes afin d'obtenir une prévision en termes bruts.

Ces modèles de type ARIMA connaissent un grand succès et de nombreux raffinements dans leur spécification (ARCIMA, ARFIMA ...) sont appliqués au domaine particulier de la finance. Il est à noter qu'on démontre que le lissage exponentiel est un cas particulier de modèle ARIMA.

Le Schéma 2 illustre la démarche de la méthodologie de Box et Jenkins.

Schéma 2 – Méthodologie de Box et Jenkins



L'examen visuel du corrélogramme est la seule information initiale dont le prévisionniste dispose pour faire son choix.

Si le corrélogramme est caractéristique d'un modèle, la décision est immédiate ; en réalité, le plus souvent, on est confronté à plusieurs possibilités qu'il faut tester les unes après les autres. Des procédures existent en vue de définir des critères rigoureux concernant la sélection des modèles.

Le modèle étant spécifié tant sous sa forme que par son degré, l'univers des modèles possibles est donc important et la tentation d'effectuer un très grand nombre de tests existe.

Très vite l'utilisateur risque d'être noyé sous les résultats (à chaque spécification correspond une émission de prévision). Ainsi, d'un système conçu initialement pour être très automatique, on aboutit rapidement à une perte en temps humain (l'utilisateur) et machine, ceci pour une amélioration parfois faible de la qualité prévisionnelle.

La méthodologie de Box et Jenkins s'avère donc plus complexe dans sa mise en œuvre et dans sa gestion au quotidien que les techniques de lissage exponentiel pour un gain en qualité incertain. Ceci explique que cette méthode d'analyse est rarement employée en entreprise (et dans les progiciels de prévision) ; elle reste réservée aux travaux de recherche universitaires sur données

micro-économiques telles que des produits financiers (actions, produits dérivés, taux d'intérêt ...) à haute fréquence (jour ou parfois heure).

L'approche endogène, lissage exponentiel ou méthodologie de Box et Jenkins, autorise le traitement en chaîne d'un grand nombre de séries chronologiques ; en effet, l'intervention humaine est réduite au minimum, voire nulle, dans le cas d'une utilisation automatique.

Cette méthode de prévision ne peut s'appliquer qu'à des chroniques relativement stables, peu sujettes aux à-coups de la conjoncture ou aux influences de la politique marketing de la firme. Certains produits du secteur de la grande consommation correspondent à ce profil ; en revanche, les produits de type industriel, très liés à l'environnement économique, et les produits de grande consommation, où la concurrence joue un rôle important, ne peuvent être traités par cette technique.

Dans le cadre d'un système de gestion de stock, ce système d'analyse de séries permet de traiter rapidement un grand nombre de références et ainsi d'élaborer des prévisions peu coûteuses à un niveau très fin.

Ainsi une « boîte noire » (le programme informatique) traite l'historique de la série des ventes et produit une prévision à l'horizon choisi par l'utilisateur ; celui-ci pouvant être ignorant du fonctionnement du système.

Cet avantage a une contrepartie évidente : par la philosophie même de cette approche, le retournement de conjoncture s'avère difficile à prévoir. En outre, l'impact de la conjoncture ou des dépenses publicitaires ne peuvent être intégrés efficacement. Le système ne peut en tenir compte qu'*a posteriori* par la rectification de certains paramètres, mais en aucun cas il ne peut anticiper. Pour le prévisionniste ceci constitue un handicap important.

Les méthodes exogènes apportent des éléments de réponse au problème de l'anticipation des retournements de conjoncture et de l'intégration des variables dont on sait pertinemment qu'elles influent sur la série à prévoir.

2. L'approche exogène

La recherche d'une explication des fluctuations d'une chronique à l'aide d'autres séries sert de cadre général aux méthodes exogènes. Dans cette partie, sont décrits les aspects méthodologiques fondamentaux²² ; la deuxième partie (approche sectorielle) de cet ouvrage indique les utilisations possibles de ces techniques, à partir d'exemples pratiques.

Nombre de variables économiques sont liées entre elles. L'exemple le plus classique est celui de la relation existant entre consommation et revenu. Nous ne retiendrons pas d'exemples de liaison macro-économique : ce n'est pas l'objet de cet ouvrage. Mais, à l'intérieur d'un secteur d'activité, d'un marché ou d'une firme, des relations peuvent également être mises en évidence. Les exemples ne manquent pas, comme nous le verrons dans les chapitres consacrés aux applications sectorielles : ventes expliquées par des logements demandés, des taux d'intérêt, des dépenses publicitaires, des variables climatiques, des promotions ... Il s'agit de l'ensemble des données suivies par les économistes d'entreprise, les prévisionnistes et les chefs de produit.

Le but de ce qui suit est de présenter les méthodes de régression simple et multiple qui permettent :

- d'obtenir la quantification, aussi précise que possible, de la liaison entre la série que l'on cherche à prévoir et la série ou les séries explicatives,
- de déterminer la précision de cette quantification afin de savoir quelle confiance accorder aux facteurs explicatifs.

²² Le lecteur intéressé par les développements de l'économétrie peut se référer à Bourbonnais (2000).

2.1. La notion d'ajustement et le modèle de régression simple

Nous avons vu que le coefficient de corrélation donne le degré de liaison entre deux chroniques. Une fois ce lien mis en évidence, nous pouvons quantifier la relation entre les deux variables économiques (les ventes et le facteur explicatif des ventes).

La formulation d'un modèle simple est alors possible sous la forme : $V_t = a_1 x_t + a_0 + e_t$

a_1 et a_0 sont les coefficients inconnus,

V_t est la série des ventes ou variable à expliquer,

x_t est la variable explicative ou variable exogène,

e_t est un terme aléatoire qui représente tout ce qui n'est pas expliqué par le modèle (la variable explicative). e_t est gaussien c'est-à-dire que les hypothèses habituelles sont respectées : espérance mathématique de l'erreur nulle ($E(e_t) = 0$), variance de l'erreur constante ($V(e_t) = \sigma_e^2$) et indépendance des erreurs ($E(e_t, e_s) = 0$, pour $t \neq s$).

Nous disposons de n observations.

Le principe de l'ajustement par les moindres carrés (méthode de régression) consiste à choisir comme estimateur pour a_0 et a_1 celui qui minimise la somme des carrés des écarts entre les valeurs ajustées (\hat{y}_t) à l'aide de la variable explicative x_t , et les observations réelles y_t :

$$\text{Min} \left(\sum_{t=1}^n e^2 \right) = \text{Min} \left(\sum_{t=1}^n (y_t - \hat{y}_t)^2 \right)$$

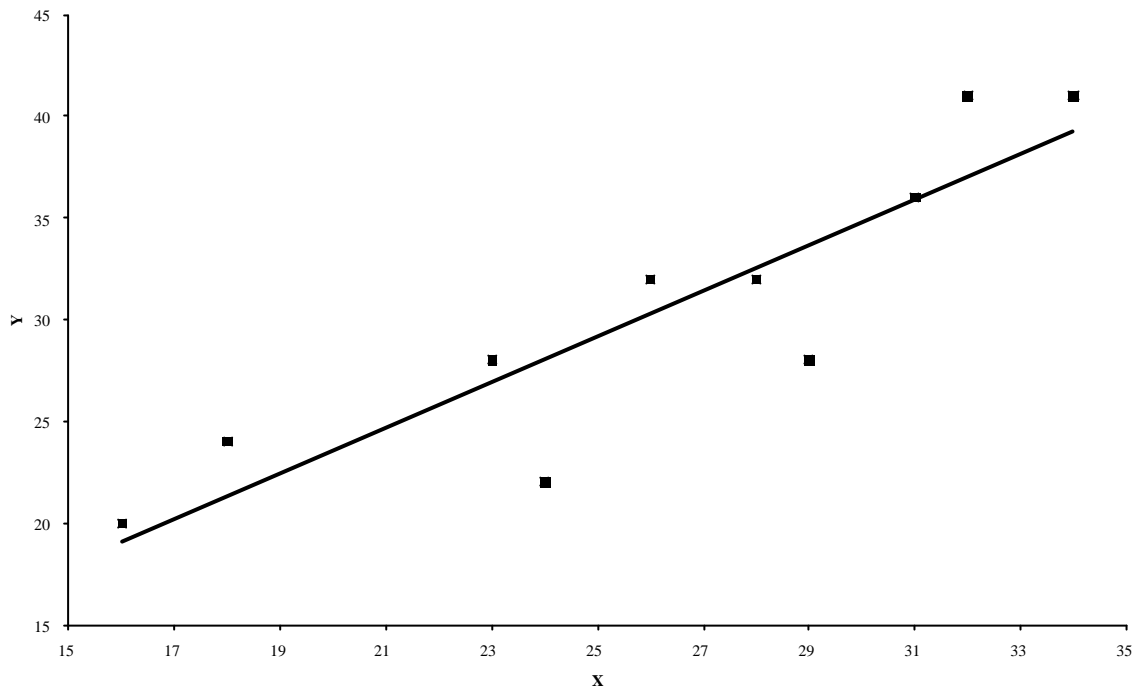
Le minimum est donné par :

$$\hat{a}_1 = \frac{\sum_{t=1}^n (y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}$$

Le Graphique 20 illustre la notion d'ajustement à partir des données du Tableau 15.

Graphique 20 – La notion d'ajustement



L'estimation de a_1 et a_0 est entachée d'une erreur (mesurée par un écart-type). Il faut déterminer si cette erreur d'estimation n'est pas trop importante et ne remet pas en cause le bien-fondé de la relation. Un test, dont le fonctionnement sera détaillé à la partie 2.3, permet de connaître la significativité d'un coefficient par rapport à zéro (un coefficient nul implique l'absence d'influence d'une variable explicative).

Ce modèle à une seule variable explicative se révèle le plus souvent trop simpliste pour rendre compte d'une réalité complexe. Il est alors souhaitable d'intégrer à ce modèle d'autres variables explicatives.

2.2. La modélisation à plusieurs variables : l'économétrie

La science économique fait souvent appel à des modèles contenant plusieurs variables exogènes. Les méthodes économétriques servent alors à estimer l'ensemble des paramètres liant la variable à expliquer aux variables explicatives. Par exemple :

$$y_t = a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + a_0 + \mathbf{e}_t$$

est un modèle à trois variables explicatives x_{1t} , x_{2t} , x_{3t} et quatre coefficients : a_1 , a_2 , a_3 , a_0 .

L'erreur \mathbf{e}_t représente l'écart entre le modèle théorique représentatif du phénomène et les valeurs observées, sur l'ensemble de la population ; \mathbf{e}_t est donc inconnue. L'estimation de l'erreur \mathbf{e}_t (e_t) est appelée le résidu (différence entre les valeurs des ventes estimées (\hat{y}_t) à l'aide des variables explicatives et les observations y_t).

Le modèle estimé s'écrit de la manière suivante :

$$y_t = \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \hat{a}_3 x_{3t} + \hat{a}_0 + e_t$$

les valeurs estimées sont donc \hat{a}_1 , \hat{a}_2 , \hat{a}_3 , \hat{a}_0 , avec $e_t =$ résidu, c'est-à-dire la différence entre la valeur observée et la valeur ajustée : $e_t = y_t - \hat{y}_t$.

$$\text{avec } \hat{y}_t = \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \hat{a}_3 x_{3t} + \hat{a}_0$$

La méthodologie d'estimation des paramètres est identique à celle d'une régression à une seule variable exogène (minimisation de la somme des carrés des écarts) ; mais la formulation devient plus complexe : nous ne pouvons plus nous situer dans un espace à deux dimensions, mais à $k + 1$ dimensions ($k =$ nombres de variables explicatives). Les tableurs permettent à l'aide de fonctions de calculer ces statistiques. Nous en verrons l'usage en 2.4.

Les modèles peuvent être synchrones : les variables sont toutes liées à la même période. Cependant, dans la pratique, lorsque nous raisonnons à partir de données à périodicité mensuelle voire hebdomadaire, l'influence des variables peut être décalée dans le temps. L'influence de la publicité ou d'une promotion peut s'exercer à 2 ou 3 mois selon le délai de réaction. Afin d'affiner les relations entre les variables, l'analyse des décalages temporels permet de se rapprocher de la réalité. Nous spécifions un modèle du type :

$$y_t = a_1 x_{1,t-\mathbf{q}_1} + a_2 x_{2,t-\mathbf{q}_2} + a_3 x_{3,t-\mathbf{q}_3} + a_0$$

où les \mathbf{q} représentent les décalages temporels entre chacune des variables $x_{i,t}$ et la série y_t .

L'utilisation de modèles à décalage permet, d'une part, une description beaucoup plus fine des mécanismes de causalité en quantifiant – précisément et de manière impartiale – les effets des facteurs explicatifs et, d'autre part, l'anticipation du retournement de conjoncture grâce à l'emploi d'indicateurs en avance. Pour le prévisionniste, ceci est d'un intérêt remarquable.

2.3. Interprétation statistique d'un modèle

Lorsque le prévisionniste a estimé un modèle économétrique après l'avoir formulé, il est important qu'il puisse juger de sa validité statistique. Pour cela, il dispose de différents tests, représentés par des statistiques calculées lors de l'estimation du modèle.

Dans un but didactique, nous partons de la présentation standard d'une équation de régression en indiquant pour chaque paramètre la manière d'interpréter sa valeur.

Soit l'équation de régression :

$$y_t = \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \hat{a}_3 x_{3t} + \hat{a}_0 + e_t$$

$(t_{\hat{a}_1}) \quad (t_{\hat{a}_2}) \quad (t_{\hat{a}_3}) \quad (t_{\hat{a}_0})$

DW

R^2

n

Ce modèle est limité à trois variables exogènes ; le développement ci-dessous reste identique dans le cas d'un nombre différent.

$\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{a}_0$, valeurs estimées des coefficients du modèle,

$t_{\hat{a}_1}, t_{\hat{a}_2}, t_{\hat{a}_3}, t_{\hat{a}_0}$ ratios de Student,

DW statistique de Durbin et Watson,

R^2 coefficient de détermination ($R =$ coefficient de corrélation multiple),

n nombre d'observations.

– Les coefficients du modèle ($\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{a}_0$)

Il s'agit des poids relatifs accordés à chacun des facteurs explicatifs. Ce sont des estimations des coefficients réels et inconnus puisque nous ne disposons que d'un échantillon de chaque chronique et non de la population dans son ensemble. Ces coefficients sont donc des variables aléatoires

assorties d'un écart-type. Ils représentent l'impact des variables explicatives sur le phénomène que l'on cherche à expliquer (les ventes).

Dans le cas d'un modèle spécifié de façon légèrement différente (introduction d'une variable explicative supplémentaire, ou remplacement d'une variable par une autre), la nouvelle estimation des coefficients de régression ($\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{a}_0$) est différente de la précédente, y compris pour les variables explicatives qui ont été conservées. En effet, le calcul des coefficients de régression tient compte des intercorrélations entre les variables explicatives ; ceci introduit donc des variations dans l'estimation de ces valeurs.

– *Les ratios de Student* ($t_{\hat{a}_1}, t_{\hat{a}_2}, t_{\hat{a}_3}, t_{\hat{a}_0}$)

Le coefficient de chaque variable explicative (nous verrons plus loin un exemple d'estimation) est en fait une variable aléatoire et, à ce titre, entaché d'une certaine erreur lors de son estimation. Cette erreur est mesurée par l'écart-type du coefficient. Plus l'estimation du coefficient est dispersée (ce qui se traduit par une valeur élevée de l'écart-type par rapport à la valeur du coefficient), plus l'influence de la variable explicative concernée sera douteuse. Le ratio de Student est calculé en faisant le rapport d'un coefficient de régression (\hat{a}_i) et de son écart-type ($\hat{S}_{\hat{a}_i}$). Ce ratio suit une loi de Student ; ceci va permettre de tester si le coefficient de régression est significativement différent de zéro (c'est-à-dire si la variable est bien explicative des ventes) ou s'il doit être considéré comme nul. Dans ce dernier cas, la variable explicative correspondante devra être éliminée de l'équation de régression car son influence sera considérée comme nulle.

Le test est mené de la manière suivante :

Si $t_{\hat{a}_i} > t_{lu}^{\alpha/2}$ lu dans une table de Student à $n - k - 1$ degrés de liberté (si le nombre d'observations est supérieur à 30, on peut prendre par approximation $t_{lu} = 2$) et pour un seuil de confiance α , le plus souvent égal à 5 % (si $\alpha = 0,05$ on accepte de se tromper dans 5 % des cas) le coefficient a_i de la variable x_i est significativement différent de zéro ; la variable x_i est bien explicative des ventes.

Dans le cas contraire, l'hypothèse d'un coefficient nul est acceptée, la variable x_i n'est pas explicative des ventes.

Cas particulier : en ce qui concerne le terme constant a_0 , le fait qu'il ne soit pas significativement différent de 0 n'a aucune importance. Dans la pratique, nous ne testerons donc pas sa significativité par rapport à 0.

– *Test de Durbin et Watson (DW)*

L'expérience montre que les modèles de régression ajustés à des séries chronologiques économiques manifestent parfois un certain degré de dépendance stochastique entre les valeurs successives du terme représentant les erreurs. En termes probabilistes, cela signifie que les erreurs sont autocorrélées, ou encore qu'une erreur commise en t a une influence sur l'erreur en $t + 1$. Durbin et Watson, en 1951, ont construit un test dont la fonction discriminante est la statistique DW .

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

La table de Durbin et Watson²³, à laquelle il faut se référer pour effectuer le test, nous indique la valeur des deux bornes d'un intervalle d_1, d_2 .

²³ Les tables statistiques sont en fin d'ouvrage.

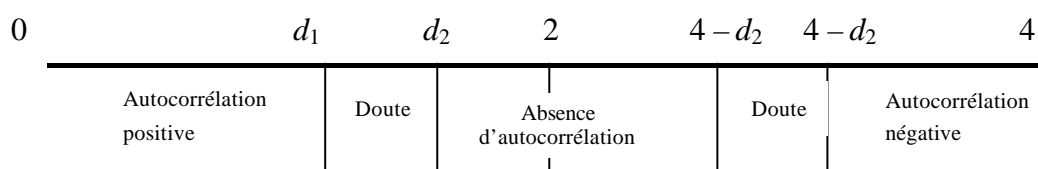
Soit DW la valeur calculée de la statistique de Durbin et Watson, la règle de décision est :

- si $DW < d_1$, on rejette l'hypothèse d'indépendance et on admet une autocorrélation des erreurs,
- si $d_1 < DW < d_2$, on est dans la zone d'indétermination des tables : il y a doute,
- si $d_2 < DW < 4 - d_2$, on accepte l'hypothèse de l'indépendance des erreurs,
- si $4 - d_2 < DW < 4 - d_1$, on est dans la zone d'indétermination des tables : il y a doute,
- si $4 - d_1 < DW$, on rejette l'hypothèse d'indépendance et on admet une corrélation négative des erreurs.

Dans la pratique le doute est interprété comme présomption favorable d'absence d'autocorrélation.

Le Schéma 3 résume ces règles de décision.

Schéma 3 – Interprétation de la statistique DW



Plusieurs causes peuvent entraîner cette autocorrélation :

- une mauvaise spécification du modèle, la relation au lieu d'être linéaire peut, par exemple, correspondre à un schéma logarithmique, ou en différences premières,
- l'absence d'une variable explicative importante,
- le lissage des données. C'est un phénomène apparenté à l'effet Slutsky²⁴ (création d'un cycle artificiel dans les séries). L'application du filtre moyenne mobile a pour conséquence automatique de dégrader la statistique de Durbin-Watson sans pour cela entraîner une remise en cause du modèle.

Dans la majorité des cas, le test de Durbin et Watson ne donne qu'une présomption d'autocorrélation des résidus pour un ordre un (corrélation entre e_t et e_{t-1}) sans rechercher des liaisons à des ordres supérieurs (2, 3 ou 4). L'usage de ce test reste donc d'un intérêt limité. L'examen visuel du graphique de la série des résidus ou le calcul de son corrélogramme sont souvent plus révélateurs.

– *Coefficient de détermination (R^2)*

Le coefficient de détermination mesure la qualité de l'ajustement du modèle, il est égal au rapport de la variance expliquée par les séries explicatives sur la variance totale. Le R^2 ne préjuge pas de la qualité réelle du modèle donc de la prévision qui en découle ; en effet, il est lié au nombre de facteurs explicatifs du modèle. A titre d'exemple caricatural, un modèle estimé à partir de 5 observations et 5 facteurs explicatifs a un R^2 égal à 1 (5 équations à 5 inconnues), or ce modèle n'a aucune validité statistique et fournirait des prévisions erronées. Ainsi, il est très possible d'aboutir à un modèle excellent avec un R^2 de l'ordre de 0,5. Le coefficient de détermination ne représente que la qualité de l'ajustement par rapport à la dispersion du nuage de points.

– *Exemple d'interprétation d'une régression*

²⁴ Slutsky (1937).

Soit le modèle estimé suivant :

$$y_t = 3,4 x_{1,t} + 5,8 x_{2,t} + 7,4 x_{3,t} + 3 + e_t$$

(2,1) (0,8) (4,4) (7,1)

$$n = 54$$

$$R^2 = 0,84$$

$$DW = 1,79$$

(.) = ratio de Student

La variable explicative $x_{2,t}$ n'est pas significative, son ratio de Student (0,8) étant très inférieur à 2, en revanche, les variables $x_{1,t}$ et $x_{3,t}$ sont significatives. La variable $x_{2,t}$ doit donc être éliminée de l'équation de régression, puis le modèle réestimé.

La nouvelle estimation de ce modèle après avoir retiré la variable $x_{2,t}$ est la suivante :

$$y_t = 2,6 x_{1,t} + 6,5 x_{3,t} + 4,12 + e_t$$

(2,8) (7,5) (1,4)

$$n = 54$$

$$R^2 = 0,81$$

$$DW = 1,88$$

(.) = ratio de Student

A la lecture des résultats, nous observons que :

- le R^2 a légèrement diminué puisqu'un facteur explicatif a été retiré ;
- les valeurs des coefficients sont modifiées par rapport à la première estimation ;
- les ratios de Student des variables explicatives ont tous augmenté en particulier pour $x_{3,t}$, ce qui signifie que la corrélation entre $x_{2,t}$ et $x_{3,t}$ est importante (elles expliquent le même phénomène, c'est l'effet de masque). Le fait d'avoir retiré la variable $x_{2,t}$ améliore considérablement l'estimation du coefficient de $x_{3,t}$;
- le ratio de Student du terme constant est inférieur à 2, le coefficient a_0 n'est donc pas significativement différent de 0, cela n'a pas d'importance ;
- le DW de Durbin et Watson ne laisse pas présager d'une autocorrélation des résidus ($k = 2$, la lecture de la table de Durbin et Watson indique : $d_1 = 1,49$; $d_2 = 1,64$), ce qui est favorable quant à la spécification du modèle ;
- le coefficient de détermination R^2 traduit un ajustement correct en regard du nombre d'observations.

En conclusion, ce modèle s'avère maintenant satisfaisant.

2.4. Exemples d'estimation économétrique à l'aide des outils sur tableur Excel

Soit un modèle économétrique composé d'une variable à expliquer (les ventes) et de trois variables explicatives (cf. Tableau 20) connues sur trois ans :

- VENTES = ventes CVS de l'entreprise en volume,
- PROMO = les dépenses en promotion en KF,
- PUB = les dépenses publicitaires en KF,
- TREND = l'évolution « naturelle²⁵ » des ventes représentée par une tendance linéaire.

²⁵ La croissance de la demande primaire.

Tableau 20 – Ventes, promotion, publicité

DATES	VENTES	PROMO	PUB.	TREND	PREVISION
Année 1-J	7281	597	3109	1	5571,27
F	9163	2139	267	2	7315,90
M	7769	859	3549	3	6350,85
A	8485	717	5083	4	6721,90
M	7606	2033	4051	5	8694,64
J	7364	1863	826	6	7413,00
J	6711	1393	4590	7	7981,23
A	3407	603	117	8	5229,43
S	4862	1702	3388	9	8286,45
O	11464	3891	2668	10	11841,74
N	10291	3968	2615	11	12048,06
D	9789	2867	1130	12	9777,90
Année 2-J	8082	250	5948	13	7068,48
F	8663	342	1715	14	5890,18
M	7720	823	366	15	6342,99
A	7859	584	6005	16	7934,60
M	6936	813	115	17	6429,19
J	6292	0	0	18	5108,52
J	7572	0	2336	19	5990,40
A	3588	1077	0	20	7118,83
S	7152	874	674	21	7096,56
O	10672	2978	5	22	10525,21
N	8522	1066	7338	23	9857,17
D	8688	30	4036	24	7084,09
Année 3-J	5772	803	5281	25	8906,04
F	7682	1984	1887	26	9853,57
M	8814	1044	115	27	7759,05
A	10864	3123	0	28	11332,27
M	5461	572	4348	29	8576,00
J	13109	4483	69	30	13844,76
J	16649	5328	5012	31	17035,95
A	12467	3890	71	32	13029,74
S	18277	6348	2128	33	17976,97
O	20836	7210	3879	34	20120,21
N	18253	3790	3782	35	14394,00
D	17483	3678	6152	36	15097,82
Année 4-J		597	3109	37	8951,53
F		2139	267	38	10696,16
M		859	3549	39	9731,11
A		717	5083	40	10102,16
M		2033	4051	41	12074,90
J		1863	826	42	10793,25
J		1393	4590	43	11361,49
A		603	117	44	8609,69
S		1702	3388	45	11666,71
O		3891	2668	46	15222,00
N		3968	2615	47	15428,32
D		2867	1130	48	13158,16

– Utilisation de la fonction « DROITREG »²⁶

²⁶ Cf. Fichier Web pour utilisation de cette fonction.

Tableau 21 – Résultat de la fonction

93,8960	0,3373	1,6922	3418
34,5297	0,14276	0,1960	731,4
0,8181	1870,1717	#N/A	#N/A
47,9904	32	#N/A	#N/A
503546068,6	111921350,4	#N/A	#N/A

Ratio de Student : 2,72 = 93,89/34,53 2,36 8,63 4,67

Tableau 22 – Tableau d'équivalence

\hat{a}_3	\hat{a}_2	\hat{a}_1	\hat{a}_0
$\hat{S}_{\hat{a}_3}$	$\hat{S}_{\hat{a}_2}$	$\hat{S}_{\hat{a}_1}$	$\hat{S}_{\hat{a}_0}$
R^2	\hat{S}_e	#N/A	#N/A
F^*	ddl	#N/A	#N/A
SCT	SCR	#N/A	#N/A

Sur la première ligne des Tableau 21 et Tableau 22, les valeurs des coefficients sont estimées (attention ! les coefficients apparaissent dans l'ordre inverse de la sélection) ; sous chaque coefficient figure son écart-type. Le coefficient de détermination $R^2 = 0,81$ est donné à l'intersection de la première colonne et de la troisième ligne. Le nombre de degrés de liberté ($ddl =$ nombre d'observations – nombre de paramètres estimés, $36 - 4 = 32$) figure à l'intersection de la colonne 2 et de la ligne 4.

Les t de Student ne sont pas fournis par la fonction mais sont calculés (rapport du coefficient sur son écart-type) sur la dernière ligne.

(Les signes #N/A signifient que le tableur Excel n'a aucune valeur à indiquer à cet endroit).

Signification de quelques statistiques supplémentaires pour les initiés :

$$\text{Ecart-type résiduel : } \hat{S}_e = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n-k-1}} = 1870,17$$

$$\text{Fisher empirique : } F^* = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}$$

$$\text{Somme des carrés totaux : } SCT = \sum_{t=1}^n (y_t - \bar{y})^2$$

$$\text{Somme des carrés des résidus : } SCR = \sum_{t=1}^n e_t^2$$

– *Utilisation de l'utilitaire d'analyse*

L'utilitaire d'analyse s'obtient à partir du menu « outil » après avoir été préalablement chargé à l'aide des macros complémentaires.

Tableau 23 – Résultat de la fonction

RAPPORT DÉTAILLÉ

<i>Statistiques de la régression</i>	
Coefficient de détermination multiple	0,90
Coefficient de détermination R ²	0,82
Coefficient de détermination R ²	0,80
Erreur-type	1870,17
Observations	36,00

	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>
Constante	3418,39	731,39	4,67
PROMO	1,69	0,20	8,63
PUB.	0,34	0,14	2,36
TREND	93,90	34,53	2,72

Nous retrouvons les mêmes résultats sous une forme différente.

N. B. : En cas de modifications de données sur le tableau initial, l'utilisation de la fonction « DROITEREG » met à jour les résultats de calcul ; en revanche, ce n'est pas le cas par l'utilisation de l'utilitaire d'analyse.

– *Interprétations statistique et économique, puis prévision*

Les ratios de Student sont tous supérieurs à 2, les coefficients des variables explicatives sont significativement différents de 0. Les trois facteurs explicatifs contribuent bien à l'explication des ventes. Ce modèle est donc valide sur le plan statistique.

Le modèle est une combinaison des trois variables : PROMO, PUB et TREND. La plus contributive à l'explication des ventes est la Promotion (*t* de Student = 8,63 le plus élevé), il s'agit aussi de la plus rentable, chaque KF investi en Promotion rapporte 1,69 unité de ventes, alors que chaque KF investi en Publicité rapporte 0,34 unité de ventes.

Connaissant les dépenses promotionnelles et publicitaires, la prévision est calculée par application du modèle :

$$\text{VENTE} = 3418,39 + 1,69 \text{ PROMO} + 0,34 \text{ PUB} + 93,90 \text{ TREND}$$

Soit pour janvier de l'année 4 :

$$\text{PREVISION}_{4-J} = 3418,39 + 1,69 \text{ PROMO}_{4-J} + 0,34 \text{ PUB}_{4-J} + 93,90 \text{ TREND}_{4-J}$$

$$\text{PREVISION}_{4-J} = 3418,39 + 1,69 \times 597 + 0,34 \times 3109 + 93,90 \times 37 = 8951,53$$

Etc.

2.5. La sélection de variables explicatives

Dans la pratique, le prévisionniste d'entreprise est souvent confronté au choix de plusieurs variables explicatives candidates pour expliquer les ventes. Or, le plus souvent ces variables apportent la même information et ne peuvent donc pas figurer dans un même modèle. Nous allons examiner quatre méthodes qui vont nous permettre de retenir les variables suivantes :

- les plus corrélées avec la variable à expliquer,
- les moins corrélées entre elles.

Nous allons illustrer chacune des méthodes à partir des données du Tableau 24.

Tableau 24 – Quelles variables retenir ?

y_t	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$
12	2	45	121
14	1	43	132
10	3	43	154
16	6	47	145
14	7	42	129
19	8	41	156
21	8	32	132
19	5	33	147
21	5	41	128
16	8	38	163
19	4	32	161
21	9	31	172
25	12	35	174
21	7	29	180

– Toutes les régressions possibles.

La méthode la plus sûre est sans aucun doute de tester toutes les équations de régression possibles et de choisir la meilleure, mais cela est pratiquement impossible au regard de l'ampleur des calculs.

En effet lorsque k variables explicatives sont prises en considération, le nombre total d'équations de régression différentes est égal à $2^k - 1$, soit par exemple, 1 023 équations pour dix variables explicatives et 1 048 575 équations pour vingt variables explicatives.

Dans notre exemple cela engendre $2^3 - 1 = 7$ modèles différents. Pour ne pas alourdir la présentation de ce chapitre, nous ne présentons pas tous les résultats. Le choix du meilleur modèle s'effectue de la manière suivante :

- par élimination des modèles dans lesquels au moins une variable explicative n'est pas significative (hormis le terme constant),
- en retenant parmi les modèles restants, celui dont le coefficient de détermination R^2 est le plus élevé.

Dans notre exemple, le lecteur pourra vérifier que c'est le modèle contenant les variables $x_{1,t}$ et $x_{2,t}$ qui est sélectionné.

– L'élimination progressive (« Backward Elimination »)

Cette procédure consiste, sur le modèle complet à k variables explicatives, à éliminer de proche en proche (c'est-à-dire en réestimant l'équation après chaque élimination) les variables explicatives dont les t de Student sont en dessous du seuil critique. Cette procédure n'est utilisable que si la première équation peut être effectivement estimée, ce qui n'est pas toujours le cas. En effet, lorsque le modèle comporte un nombre trop important de variables explicatives, le risque de colinéarité entre ces variables est élevé.

Première régression :

	$\hat{\alpha}_3$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_0$
Coefficient	-0,0371	-0,3814	0,8019	32,8913
Ecart-type	0,05	0,16	0,30	11,66
t de Student	0,71	2,44	2,69	2,82

La variable $x_{3,t}$ n'est pas significative ($t_{cal} = 0,71 < 2$), elle est donc retirée du modèle :

	\hat{a}_2	\hat{a}_1	\hat{a}_0
Coefficient	-0,3281	0,7149	25,8421
Ecart-type	0,13	0,27	6,06
<i>t de Student</i>	2,44	2,68	4,26

Toutes les variables sont maintenant significatives ($t_{cal} > 2$), le modèle est validé (les variables $x_{1,t}$ et $x_{2,t}$ sont sélectionnées).

– *Le processus d'introduction progressive (« Stepwise Regression »)*

Cette technique utilise une démarche inverse ; les variables explicatives sont sélectionnées une par une jusqu'à ce que l'équation de régression se dégrade. Dans un premier passage, la variable la plus fortement corrélée avec la série à expliquer est choisie. Puis au deuxième stade, la variable qui explique le plus la variance de la série à expliquer (selon le critère du *t de Student*) est introduite. Le processus d'introduction s'arrête lorsque plus aucune variable complémentaire ne contribue à l'explication de la variance de la série à expliquer.

- Etape 1 : Sélection de la variable explicative la plus corrélée avec la variable à prévoir y_t .
Nous calculons les trois coefficients de corrélation entre y_t et chacune des trois variables ainsi que son *t de Student* associé (cf. 1.1.1) :

	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$
Corrélation	0,72	-0,69	0,48
<i>t de Student</i>	3,88	3,61	2,05

La variable explicative $x_{1,t}$ est donc retenue (coefficient de corrélation le plus élevé).

- Etape 2 : Estimation des deux modèles avec comme première variable explicative $x_{1,t}$ puis alternativement $x_{2,t}$ et $x_{3,t}$.

$$y_t = a_1 x_{1,t} + a_2 x_{2,t} + a_0 + e_t$$

	\hat{a}_2	\hat{a}_1	\hat{a}_0
	-0,3281	0,7149	25,8421
	0,13	0,27	6,06
<i>t de Student</i> =	2,44	2,68	4,26

$$y_t = a_1 x_{1,t} + a_3 x_{3,t} + a_0 + e_t$$

	\hat{a}_3	\hat{a}_1	\hat{a}_0
	0,0232	0,9272	8,6092
	0,06	0,35	7,28
<i>t de Student</i> =	0,42	2,62	1,18

La deuxième variable sélectionnée (ici $x_{2,t}$) est celle dont le *t de Student* est le plus élevé en valeur absolue.

- Etape 3 : Estimation d'un modèle avec comme variables $x_{1,t}$ et $x_{2,t}$ (les deux premières sélectionnées) et $x_{3,t}$ (la seule qui reste).

$$y_t = a_1 x_{1,t} + a_2 x_{2,t} + a_3 x_{3,t} + a_0 + e_t$$

Nous connaissons le résultat, car il s'agit de la première estimation de la procédure de l'élimination progressive. La procédure est donc arrêtée car la variable x_{3t} n'est pas significative de par l'interprétation du t de Student.

Le modèle final est donc le même que dans les procédures précédentes.

– *La régression à étages (« Stagewise Regression »)* (cf. fichier Web).

C'est un processus de sélection de variables explicatives permettant de rechercher les contributions marginales, mais significatives, de chacune des variables explicatives candidates.

Comme pour la méthode précédente, la variable ayant le coefficient de corrélation le plus élevé avec la série à prévoir est retenue.

Nous procédons alors à une première régression, qui dégage un résidu. Ce résidu est considéré à son tour comme une nouvelle variable à expliquer. Dans un deuxième temps, la série explicative la mieux corrélée avec ce premier résidu est sélectionnée, une nouvelle régression à deux variables dégage un nouveau résidu que l'on cherchera aussi à expliquer, etc.

- Etape 1 : Sélection de la variable explicative la plus corrélée avec la variable à prévoir y_t .

Nous calculons les trois coefficients de corrélation entre y_t et chacune des trois variables ainsi que son t de Student associé. La variable explicative $x_{1,t}$ est donc retenue (coefficient de corrélation le plus élevé).

- Etape 2 : Estimation d'un modèle avec une variable explicative $x_{1,t}$ et calcul du résidu.

$$y_t = \hat{a}_1 x_{1,t} + \hat{a}_0 + e_{1,t}$$

$$e_{1,t} = y_t - (\hat{a}_1 x_{1,t} + \hat{a}_0)$$

$$y_t = 1,0118 \times x_{1,t} + 11,57 + e_{1,t}$$

Les coefficients estimés nous permettent de calculer le résidu $e_{1,t}$:

$$e_{1,t} = y_t - (1,0118 \times x_{1,t} + 11,57)$$

Par exemple :

$$e_{1,1} = 12 - (1,0118 \times 2 + 11,57) = -1,59$$

$$e_{1,2} = 14 - (1,0118 \times 1 + 11,57) = 1,42$$

...

$$e_{1,14} = 21 - (1,0118 \times 7 + 11,57) = 2,35$$

- Etape 3 : calcul des coefficients de corrélation simple entre $e_{1,t}$ et les différents $x_{i,t}$.

	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$
Corrélation	0,00	-0,53	0,10
t de Student	0,00	2,32	0,39

La deuxième variable sélectionnée (ici $x_{2,t}$) est celle dont le coefficient de corrélation est le plus élevé en valeur absolue.

On remarque que le coefficient de corrélation entre $e_{1,t}$ et $x_{1,t}$ est nul car la variable $e_{1,t}$ représente la série des ventes de laquelle nous avons retiré l'influence de $x_{1,t}$.

- Etape 4 : Estimation d'un modèle avec deux variables explicatives $x_{1,t}$ et $x_{2,t}$ puis calcul du résidu.

$$e_{2,t} = y_t - (\hat{a}_1 x_{1,t} + \hat{a}_2 x_{2,t} + \hat{a}_0)$$

Les coefficients estimés nous permettent de calculer le résidu $e_{2,t}$:

$$e_{2,t} = y_t - (0,71 \times x_{1,t} + 0,33 \times x_{2,t} + 25,84)$$

Par exemple :

$$e_{2,1} = 12 - (0,71 \times 2 + 0,33 \times 45 + 25,84) = -0,51$$

$$e_{2,2} = 14 - (0,71 \times 1 + 0,33 \times 43 + 25,84) = 1,55$$

...

$$e_{2,14} = 21 - (0,71 \times 7 + 0,33 \times 29 + 25,84)$$

- Etape 5 : calcul des coefficients de corrélation simple entre $e_{2,t}$ et les différents $x_{i,t}$.

	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$
Corrélation	0,00	0,00	-0,16
t de Student	0,00	0,00	0,60

La procédure de sélection s'arrête car plus aucun coefficient de corrélation entre le résidu et les variables explicatives n'est significativement différent de 0.

L'analyse des résidus à chaque étape permet de tirer le maximum d'informations des variables explicatives en minimisant les inter-corrélations. Economiquement, cela se justifie pleinement : une industrie peut dépendre à 60 % d'un premier secteur, de 30 % d'un autre et de 10 % d'un dernier. Ainsi l'explication résiduelle du deuxième secteur est mise en lumière après soustraction de l'explication par le premier secteur ; quant au troisième, il apparaît avec plus de netteté sur le résidu des 90 % déjà expliqués. Ainsi, nous pouvons mettre en évidence les explications marginales mais contributives de chacune des variables.

L'approche économétrique est très riche car elle recherche une explication non pas dans les variations de la chronique elle-même (comme les méthodes endogènes) mais au travers des fluctuations d'autres chroniques dont on a l'habitude d'évaluer intuitivement l'influence. *Elle permet de tester explicitement des relations établies de façon implicite.*

L'inconvénient de ce type de méthode réside dans une mise en œuvre un peu plus lourde. Cependant les tableurs permettent une grande facilité de calcul (comme nous l'avons vu dans ce chapitre). L'utilisateur ne doit intervenir que dans la spécification du modèle, l'évaluation de sa validité statistique, le suivi du modèle et des indicateurs ainsi que la recherche et la constitution des historiques de séries explicatives.

Chapitre 5

EVALUATION DES METHODES DE PREVISION

Ce dernier chapitre traite, dans une première partie, de l'évaluation de la qualité de la prévision. Cette question paraît *a priori* triviale. Elle ne l'est en fait pas du tout. Les entreprises éprouvent souvent de grandes difficultés à mesurer « objectivement » la qualité d'une prévision. La première étape de cette mesure consiste à fixer des critères qui permettent d'évaluer l'exactitude de la prévision pour un article donné (section 1) puis, pour des entreprises gérant un grand nombre de références, de proposer des méthodes d'évaluation synthétique de la qualité de la prévision (section 2). La section 3 présente des procédures d'évaluation de la qualité de la prévision.

Une préoccupation majeure avant de mettre en place un système est de choisir la « meilleure » méthode. La sélection de celle-ci a fait l'objet de nombreuses tentatives de mesure, qui se situent dans l'absolu, c'est-à-dire sans référence réelle au type de données, au secteur, et à l'horizon de prévision. La section 4 de ce chapitre décrit les tentatives de comparaison entre méthodes et met en valeur leur caractère discutable : si des éléments probants permettent de se déterminer en faveur de l'une ou de l'autre des techniques, ce n'est pas dans ce type de recherche qu'on les trouvera.

La section 5 expose comparativement les avantages et les inconvénients des principales techniques de prévision ; les jugements qui sont exposés sont de nature qualitative et ils sont marqués par la prudence. En effet, les méthodes ne sont jamais entièrement comparables parce qu'elles ne sont pas conçues rigoureusement pour des applications et des horizons semblables.

1. Mesure de la qualité d'une prévision

1.1. Les indicateurs de mesure

Il existe de très nombreux indicateurs²⁷ permettant d'évaluer la qualité de la prévision, nous présentons ici les plus importants.

– Erreur relative en pourcentage à la période t (ER_t)

Il s'exprime comme :

$$ER_t = \frac{|x_t - \hat{x}_t|}{x_t} \times 100$$

où :

x_t est la réalisation à la période t ,

\hat{x}_t est la prévision à la période t .

Il s'agit d'une approche simple et classique, qui est renouvelée chaque fois que de nouvelles réalisations connues permettent de calculer des écarts.

– Indicateur de la valeur absolue moyenne des écarts (*Mean Absolute Deviation*)

Si on veut avoir une vision synthétique qui opère sur un ensemble de T périodes de prévision, le calcul de la valeur absolue moyenne des erreurs s'impose. Il est défini à l'instant t par :

²⁷ Cf. Bourbonnais et Terraza, Chapitre 7, 1998.

$$MAD_t = \frac{\sum_{i=t-T+1}^t |EPS_i|}{T}$$

avec : $EPS_i = x_i - \hat{x}_i$

– Indicateur de la variance de l'erreur de prévision (*MSE*, « *Mean Squared Error* »)

Afin de rendre compte des erreurs importantes, tout en ignorant, comme dans la formule précédente, leur signe, l'indicateur utilisé communément est celui de la variance de l'erreur de prévision. Il est défini à l'instant t par :

$$MSE_t = \frac{\sum_{i=t-T+1}^t EPS_i^2}{T}$$

La racine carrée de cet indicateur \sqrt{MSE} est similaire à l'écart-type de l'erreur de prévision non centrée sur sa moyenne. Sa valeur dimensionne donc le stock de sécurité²⁸ pour l'article considéré.

– Indicateur de comparaison des performances de deux prévisions, statistique dite « U » de Theil²⁹.

Cet indicateur compare la prévision de l'entreprise avec une méthode de prévision « naïve » consistant à prendre comme nouvelle prévision la dernière réalisation. Il est défini à l'instant t par :

$$U = \sqrt{\frac{\sum_{i=1}^{i=T-1} (CRP_{i+1} - CRR_{i+1})^2}{\sum_{i=1}^{i=T-1} (CRR_{i+1})^2}}$$

Avec :

T = nombre de périodes de prévision (exemple = 12)

CRP = Changement Relatif Prévu (exemple = + 5,2 %)

CRR = Changement Relatif Réel (exemple = + 3,2 %)

L'interprétation des valeurs de la statistique U peut se résumer de la façon suivante :

$U = 1 \Rightarrow$ la technique de prévision est équivalente à une méthode naïve qui consiste à prendre comme prévision la dernière réalisation.

$U < 1 \Rightarrow$ la technique de prévision est meilleure que cette méthode et d'autant meilleure que U est proche de 0.

$U > 1 \Rightarrow$ la technique de prévision est très médiocre, car elle est pire que cette méthode.

1.2. La prévision est-elle biaisée ?

Une autre approche de la mesure de la qualité de la prévision consiste à détecter l'existence d'un biais. Le biais de prévision doit être détecté rapidement car il est générateur, à terme, soit d'un surstock soit d'une rupture. Hormis l'indicateur *AWS*, présenté au chapitre 3, qui permet de détecter des ruptures de tendance et donc des biais, nous présentons deux tests statistiques supplémentaires.

La première procédure consiste à spécifier, puis à estimer le modèle de régression suivant :

$$x_t = a_0 + a_1 \hat{x}_t$$

où :

²⁸ Cf. Bourbonnais et Vallin, 1995.

²⁹ Theil, 1966.

x_t est la valeur observée en t ,

\hat{x}_t est la valeur prévue en t .

Si le coefficient a_0 est significativement différent de 0, cela révèle l'existence d'un biais. Le coefficient a_0 devrait en effet être nul (non significativement différent de 0) si en moyenne les prévisions sont égales aux réalisations.

L'exemple suivant illustre cette méthode. Soit le modèle estimé :

$$x_t = 3,32 + 0,931 \hat{x}_t$$

(0,42) (9,52)

$$R^2 = 0,91$$

$T = 18$ (nombre de périodes de prévision)

(.) = t de Student

Le coefficient a_0 n'est pas significativement différent de 0 puisque $t^* = 0,42 < t_{16}^{0,05} = 2,13$, les prévisions ne sont donc pas biaisées.

Le deuxième test consiste à calculer le rapport de la racine de l'indicateur MSE à la moyenne arithmétique des erreurs de prévision sur T périodes : $\frac{\sqrt{MSE_t}}{EPS}$

$$\text{Avec : } \overline{EPS} = \frac{\sum_{i=t-T+1}^t EPS}{T}$$

Puis à comparer cet indicateur à la racine carrée de T divisée par 2, soit le test suivant :

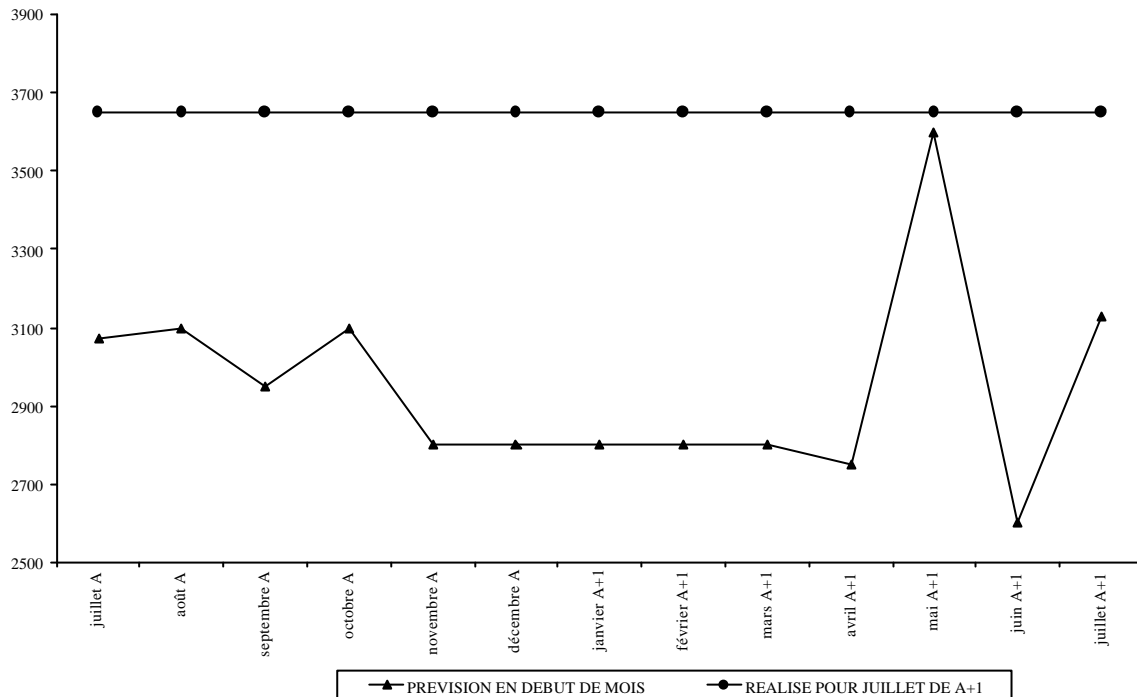
Si $\frac{\sqrt{MSE_t}}{EPS} \geq \frac{\sqrt{T}}{2}$, alors il n'y a pas de présomption de biais.

1.3. Convergence de la prévision

La question peut se poser de l'amélioration de la qualité de la prévision lorsque l'on se rapproche de la date de réalisation de la vente. En théorie, il semble plus facile d'élaborer une prévision pour le mois prochain que pour dans 12 mois. Les prévisions glissantes font que, pour un même mois de réalisation, des prévisions différentes s'accumulent : celle calculée il y a 12 mois, puis celle calculée il y a 11 mois, etc., jusqu'à celle calculée le mois d'avant ; ainsi au bout de 12 mois, on dispose de 12 prévisions – émises successivement – différentes pour un même mois. Il peut être intéressant d'analyser l'évolution de la prévision au fur et à mesure que l'on s'approche de la réalisation. Une prévision très fluctuante d'une émission à l'autre est perturbante pour la planification de la production, car elle oblige à des révisions importantes du programme de fabrication (besoins humains et en matériel). De plus, nous pouvons nous interroger sur la convergence de la prévision, c'est-à-dire sur son amélioration au fur et à mesure que l'échéance approche.

Le Graphique 21 illustre cette convergence à partir d'un exemple réel issu d'un constructeur automobile (prévision d'immatriculations d'un modèle automobile). L'interprétation est la suivante : la droite représente la réalisation pour juillet de l'année A+1, (le nombre de modèles réellement immatriculés est de 3680), la courbe représente les différentes prévisions calculées pour le mois de juillet de l'année A+1, celle calculée au 1^{er} juillet de l'année A+1 est de 3200, celle calculée en juin de l'année A+1 est de 2600, etc., en remontant jusqu'en juillet de l'année A (la prévision pour juillet de l'année A+1 est alors de 3070).

Graphique 21 – Evolution de la prévision glissante



Nous constatons que les différentes prévisions calculées depuis juillet de l'année A jusqu'en avril de l'année A+1 se sont plutôt dégradées au cours du temps. Celle calculée en mai est très proche de la réalisation, puis s'en éloigne fortement le mois suivant (juin) ; enfin, en début juillet, la prévision retend vers la valeur finale. Nous pouvons remarquer :

- que la prévision de juillet de l'année A+1 n'est pas notablement de meilleure qualité que celle de juillet de l'année A ;
- qu'après avoir été très proche de la réalisation, le prévisionniste a dégradé cette prévision ;
- qu'enfin l'amplitude des variations de la prévision sur les quatre derniers mois est très forte.

Ce type de phénomène se produit assez souvent en entreprise où nous pouvons observer que la proximité de la date de réalisation rend fébrile le prévisionniste qui, en perdant toute objectivité d'une vision de moyen terme, dégrade la prévision initiale élaborée avec plus de recul.

1.4. Détection de la cause de l'erreur

Plusieurs auteurs (Theil, 1971 ; Jorgenson, 1970 ; Granger et Newbold, 1973) proposent une mesure de l'importance relative de chaque source de l'erreur quadratique moyenne (biais, amplitude, composante strictement aléatoire).

$$\text{On pose : } E = (\bar{x} - \bar{\hat{x}})^2 + \mathbf{s}_x^2 + \mathbf{s}_{\hat{x}}^2 - 2r\mathbf{s}_{\hat{x}}\mathbf{s}_x$$

où :

\bar{x} = moyenne des réalisations,

$\bar{\hat{x}}$ = moyenne des prévisions,

\mathbf{s}_x^2 = variance des réalisations,

$\mathbf{s}_{\hat{x}}^2$ = variance des prévisions,

r = coefficient de corrélation entre valeurs réalisées et prévues.

Les conditions pour que l'erreur quadratique de prévision soit minimum sont les suivantes :

$\bar{x} = \bar{\hat{x}}$ (prévision sans biais) et $r\mathbf{s}_x = \mathbf{s}_{\hat{x}}$,

si $r = 1$, l'écart-type des prévisions et l'écart-type des réalisations sont identiques.

Nous pouvons écrire cette même équation de la manière suivante :

$$E = (\bar{x} - \bar{\hat{x}})^2 + (\mathbf{s}_{\hat{x}} - r\mathbf{s}_x)^2 + (1 - r^2)\mathbf{s}_x^2$$
$$E = EB + EA + ER$$

avec :

$EB = (\bar{x} - \bar{\hat{x}})^2$ = composante associée au biais,

$EA = (\mathbf{s}_{\hat{x}} - r\mathbf{s}_x)^2$ = composante associée à l'amplitude de l'erreur,

$ER = (1 - r^2)\mathbf{s}_x^2$ = composante aléatoire.

L'objectif du suivi des écarts du système de prévision est alors d'éliminer EB (le biais), puis ensuite de réduire EA (l'amplitude de l'erreur), en améliorant progressivement le modèle.

1.5. Quelle erreur de prévision est admissible ?

Les entreprises se posent cette question fréquemment, elles la formulent parfois autrement :

Une erreur de prévision de 15 % est-elle admissible ?

Comment déterminer un objectif de qualité de la prévision ?

Comment étalonner *a priori* les performances attendues ?

La qualité attendue de la prévision dépend étroitement du secteur d'activité et de la difficulté intrinsèque (variance de la série historique) de prévoir l'article. Par exemple, pour un produit de grande consommation, une erreur de prévision de 5% est considérée comme importante, alors que dans le domaine industriel la même erreur de prévision est jugée tout à fait acceptable.

Nous présentons une technique fondée sur le coefficient de variation (cf. chapitre 2) permettant de répondre à cette question.

Le Coefficient de Variation (CV) de la série brute rend compte de la difficulté prévisionnelle d'un historique. Plus il est élevé, plus la variance de la série est importante rapportée à la moyenne.

Cependant, une saisonnalité marquée peut augmenter artificiellement le CV. C'est pourquoi, il convient de raisonner sur la série CVS. La procédure est la suivante :

– estimation de la série CVS,

– calcul du coefficient de variation (CV_{CVS}) de la série CVS.

Si, par exemple, $CV_{CVS} = 0,12$ la plage de variation intrinsèque de cet article est de $\pm 12\%$, nous pouvons considérer que la valeur ajoutée du modèle de prévision et du prévisionniste permet de diviser par 2 cette variation, l'objectif de performance pour cet article est donc de 6% d'erreur de prévision.

Il s'agit d'un étalonnage très simple à mettre en œuvre et qui permet aux entreprises de se fixer *a priori* un objectif de qualité de prévision.

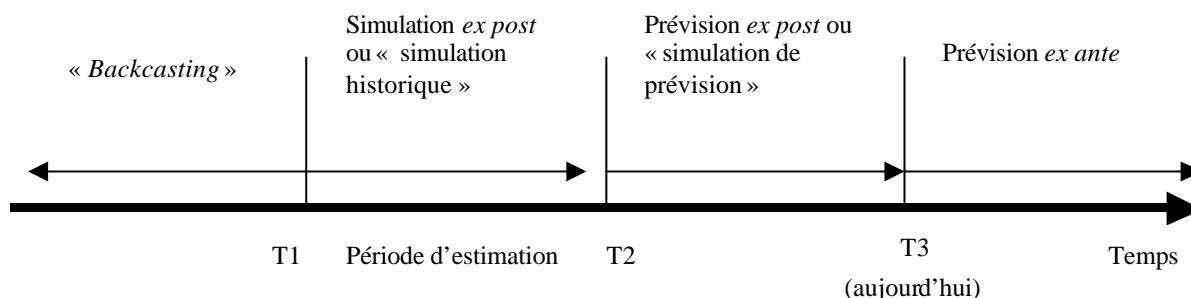
1.6. La simulation

Une autre manière permettant d'évaluer la qualité attendue des prévisions consiste à opérer des simulations des prévisions. Il convient d'être extrêmement prudent quant à leurs interprétations et les conséquences que l'on peut en tirer. En effet, cela dépend du type de simulation effectuée.

– *Les différents types de simulation*

Différents types de simulation sont possibles (cf. Schéma 4).

Schéma 4 – Différents types de simulation



Dans le Schéma 4, T1 et T2 représentent les dates de début et de fin de la période sur laquelle est estimé le modèle, T3 représente la date actuelle.

La première modalité de simulation est appelée simulation *ex post* ou simulation historique : elle correspond aux modèles de simulation. La totalité de l'historique de T1 en T2 est utilisée en ce qui concerne la série des ventes que l'on cherche à prévoir et éventuellement aussi les facteurs explicatifs. Après l'estimation des paramètres du modèle, les prévisions sont simulées sur le même historique. Ce type de simulation donne peu d'indications sur la qualité réelle attendue des prévisions³⁰. En effet, l'objectif statistique de l'estimation du modèle étant de minimiser la somme des erreurs de prévision, cela permet de vérifier seulement que les algorithmes utilisés fonctionnent convenablement.

Le deuxième type de simulation (prévision *ex post* ou simulation de prévision) implique une simulation du modèle après la période d'estimation ; une partie de l'historique connue n'est pas utilisée pour estimer les paramètres du modèle de prévision mais sert exclusivement à comparer les prévisions issues de cette estimation et les réalisations. Cette procédure permet de qualifier les résultats attendus en terme de qualité et constitue un test réellement interprétable.

Quant au « *backcasting* », il s'agit d'un type de simulation où l'on prévoit « à reculons » dans le passé : à partir d'un modèle estimé sur la période T1 à T2, on prévoit les années antérieures. L'intérêt de cette procédure, relativement peu fréquente en pratique, n'est pourtant pas mince : elle permet de tester la stabilité dynamique du modèle comme s'il évoluait vers le passé, au lieu d'aller en avant dans le temps.

Enfin, la prévision *ex ante* consiste à effectuer une prévision pour une période non connue.

2. Evaluation synthétique de la qualité de la prévision

Les formules développées aux parties précédentes s'appliquent pour un article donné. Or, pour des entreprises qui gèrent un grand nombre de références, il devient très difficile d'appréhender dans sa globalité la qualité des prévisions.

De plus, l'évaluation de l'erreur en terme de pourcentage n'a pas toujours beaucoup de sens ; en effet, il est plus grave de se tromper de 5% pour un article dont les quantités vendues sont en moyenne de 100 000 unités par an que de faire une erreur de prévision de 50% pour un article vendu à 50 unités par an (pour des valeurs unitaires comparables).

2.1. Calcul d'un indice pondéré

Pour pallier ces inconvénients, nous proposons une première technique qui permet, à la lecture d'un simple chiffre, d'avoir une vision synthétique de la qualité de la prévision.

³⁰ Pourtant c'est une pratique courante et trompeuse ...

Le principe consiste à calculer la somme des produits des erreurs de prévision pondérées par le volume ou le chiffre d'affaires.

Le Tableau 25 illustre la méthode de calcul.

Tableau 25 – Calcul d'un indice pondéré par le volume et le chiffre d'affaires

Prévision Ligne 1	REF1, 3450	REF2, 869	REF3, 900	REF4, 4309	REF5, 3300	REF6, 2100	REF7, 70	REF8, 320
Réalisation Ligne 2	3477	858	919	5074	2774	2919	134	200
Ecart Ligne 3	-0,78%	1,28%	-2,07%	-15,08%	18,96%	-28,06%	-47,76%	60,00%
Ecart absolu Ligne 4	0,78%	1,28%	2,07%	15,08%	18,96%	28,06%	47,76%	60,00%
Chiffre d'affaires Ligne 5	14000	16000	20000	15000	8000	17000	15000	22000
Chiffre d'affaires pondéré Ligne 6	109	205	413	2262	1517	4770	7164	13200
Volume Pondéré Ligne 7	27	11	19	765	526	819	64	120

En colonne figurent les références gérées (au nombre de huit), en ligne nous avons les différentes étapes de calcul. Les cinq premières lignes ne posent aucun problème de compréhension. Nous explicitons seulement les deux dernières lignes :

Chiffre d'affaires pondéré (ligne 6) = chiffre d'affaires (ligne 5) × écart absolu (ligne 4)

Volume pondéré (ligne 7) = réalisation (ligne 2) × écart absolu (ligne 4)

L'indice de la qualité de la prévision pondéré par le chiffre d'affaires est alors égal à la somme de la ligne 6 divisée par la somme de la ligne 5, soit $= 29\ 640 / 127\ 000 = 0,2334$. L'erreur de prévision moyenne tenant compte du chiffre d'affaires de chaque article est de 23,34%.

L'indice de la qualité de la prévision pondéré par le volume³¹ est alors égal à la somme de la ligne 7 divisée par la somme de la ligne 2, soit $= 2\ 351 / 16\ 355 = 0,1437$. L'erreur de prévision moyenne tenant compte du volume de chaque article est de 23,34%.

L'erreur de prévision pondérée par le chiffre d'affaires est de 23,34% et l'erreur de prévision pondérée par le volume est de 14,37%, les prévisions sont donc de moins bonne qualité pour les articles à valeur unitaire élevée.

Ces indices, calculés tous les mois, permettent de suivre de façon dynamique l'amélioration ou la détérioration de la prévision.

2.2. Evaluation graphique

Une autre approche permet, à la lecture d'un graphique, d'avoir une vision synthétique de la qualité de la prévision.

Le Tableau 26 illustre la construction du graphique pour une société qui gère 81 références :

– classement des articles par ordre croissant de l'erreur relative de prévision (en %) en valeur absolue (colonne 4),

³¹ Si cela a un sens de comparer les volumes des articles.

- calcul du poids relatif de l'article rapporté à la totalité des ventes de l'entreprise (réalisation en volume de l'article divisée par la totalité du volume des ventes),
- cumul de ces poids (en % ; colonne 6) .

Tableau 26 – Construction d'un graphique d'évaluation

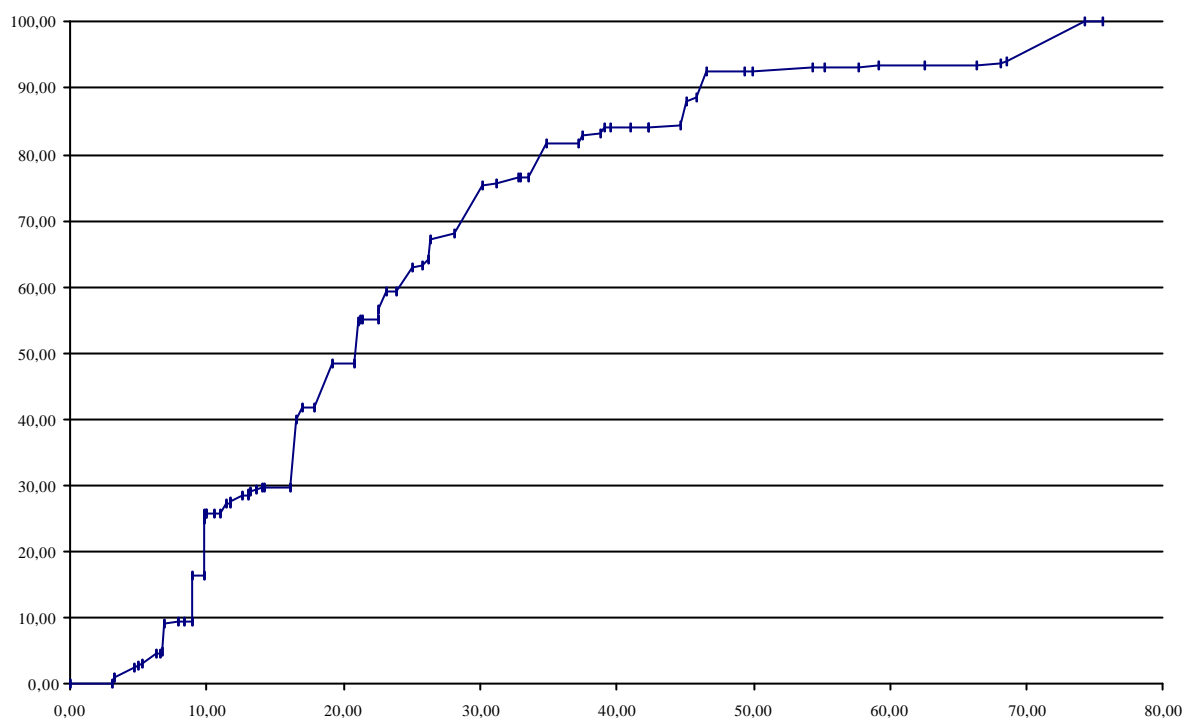
Référence Colonne 1	Réalisation Colonne 2	Prévision Colonne 3	% Ecart Colonne 4	Cumul des ventes Colonne 5	% ventes totales Colonne 6
Ref1	40	40	0,00	40	0,01
Ref2	33	34	3,03	73	0,01
Ref3	5872	5685	3,18	5945	0,86
Ref4	91	88	3,30	6036	0,87
Ref5	11021	10501	4,72	17057	2,46
Ref6	1684	1599	5,05	18741	2,70
Ref7
Ref8
Ref72	4148	6404	54,39	645488	93,05
Ref73	38	17	55,26	645526	93,06
Ref74	687	291	57,64	646213	93,16
Ref75	1853	2950	59,20	648066	93,43
Ref76	48	78	62,50	648114	93,43
Ref77	800	1331	66,38	648914	93,55
Ref78	1513	2544	68,14	650427	93,77
Ref79	1297	2186	68,54	651724	93,95
Ref80	41770	72759	74,19	693494	99,97
Ref81	180	44	75,56	693674	100,00

Le Graphique 22 représente en abscisse les erreurs en pourcentage (colonne 4) et en ordonnée (colonne 6) le cumul des poids (jusqu'à 100% du volume traité).

L'interprétation est alors aisée :

- plus la courbe se situe à gauche et croît rapidement (forme convexe), meilleure est la qualité de la prévision,
- pour juger de la qualité globale, il suffit de se placer par exemple à 10 et à 20 sur l'abscisse. Cela correspond à environ 25 % puis 49 % du volume traité. Pour 25 % du volume prévu, l'erreur de prévision est inférieure ou égale à 10 % et, pour 49 % (la moitié) du volume prévu, l'erreur de prévision est inférieure ou égale à 20%.

Graphique 22 – Evaluation synthétique d'un système de prévision



Cette technique est utile pour éviter de donner trop d'importance à une erreur isolée, même très importante. En outre elle peut aussi bien s'appliquer à des données en volume (unités), comme dans notre exemple, qu'à des données en valeur, lorsque les articles ont des unités de mesure hétérogènes.

Dans cet exemple, nous avons considéré les erreurs de prévision en valeur absolue en faisant l'hypothèse qu'il est aussi grave de se tromper en positif qu'en négatif. Cependant, certaines entreprises craignent avant tout la rupture et considèrent donc qu'une erreur de prévision positive (réalisation > prévision) est plus grave que l'inverse. La constitution de deux graphiques – l'un pour les erreurs de prévisions positives et l'autre pour les négatives – permet de répondre à cette préoccupation.

3. Procédures d'évaluation

3.1. Durée de la mesure

Une prévision ne se juge pas dès le premier essai. Que le résultat soit favorable ou défavorable, il est nécessaire de faire un essai suffisamment long pour que des écarts faibles ou forts dépassent le stade de la coïncidence pour atteindre celui de la signification statistique d'un véritable degré de précision.

Elle ne doit pas non plus être trop longue : si le long terme est nécessaire pour évaluer l'exactitude de la prévision à court terme, quand l'utilisera-t-on ?

Une règle empirique est de faire un test sur une période qui représente trois fois l'horizon de prévision. Par exemple, pour une prévision semestrielle émise tous les 3 mois, on calcule les écarts pendant une période de 18 mois, ce qui revient à tester six émissions de prévision.

Partant d'un exemple (prévision à 6 mois, actualisée tous les trimestres, en valeurs mensuelles), où le problème est une mesure de l'écart prévisions – réalisations et non une compétition entre différentes méthodes, nous proposons la démarche suivante :

Tous les 3 mois, on procède à une évaluation de la qualité de la prévision au niveau individuel (cf. 1) et synthétique (cf. 2), selon un pas mensuel et trimestriel (une prévision non fiable au niveau mensuel peut être fiable trimestriellement).

Il faut se méfier d'interpréter trop rapidement un résultat favorable. Une série brute dont le coefficient de variation de 20 % (écart-type exprimé en pourcentage de la valeur moyenne), qui une fois désaisonnalisée n'a plus qu'un coefficient de variation de 8 %, est prévue avec une erreur quadratique moyenne de 7 %, la performance prévisionnelle ne doit pas être considérée comme très satisfaisante. L'apport d'une méthode de prévision, sa précision, ne se mesurent pas de façon linéaire. Certaines composantes peuvent être faciles à dégager et donc à prévoir, la saisonnalité par exemple. Dans le cas évoqué précédemment, la précision hors saisonnalité était en réalité très faible.

En outre, la qualité ne se mesure pas de façon continue. Prenons l'exemple suivant : la prévision est erronée, à la baisse sur un marché où une reprise technique semble se faire jour. Durant les trois mois qui suivent, elle redevient précise, cependant que les chiffres réels diminuent, accréditant ainsi l'idée que cette reprise était un leurre. On peut dire alors que son apport a été important car elle a jeté un doute sur la durabilité d'une modification de tendance et elle a joué un rôle modérateur.

3.2. Comparaisons de prévisions

Plusieurs possibilités existent : elles témoignent du contexte dans lequel s'effectue l'étude, suivant que l'on essaie de prévoir la réalité ou de comparer différentes méthodes. La référence peut être :

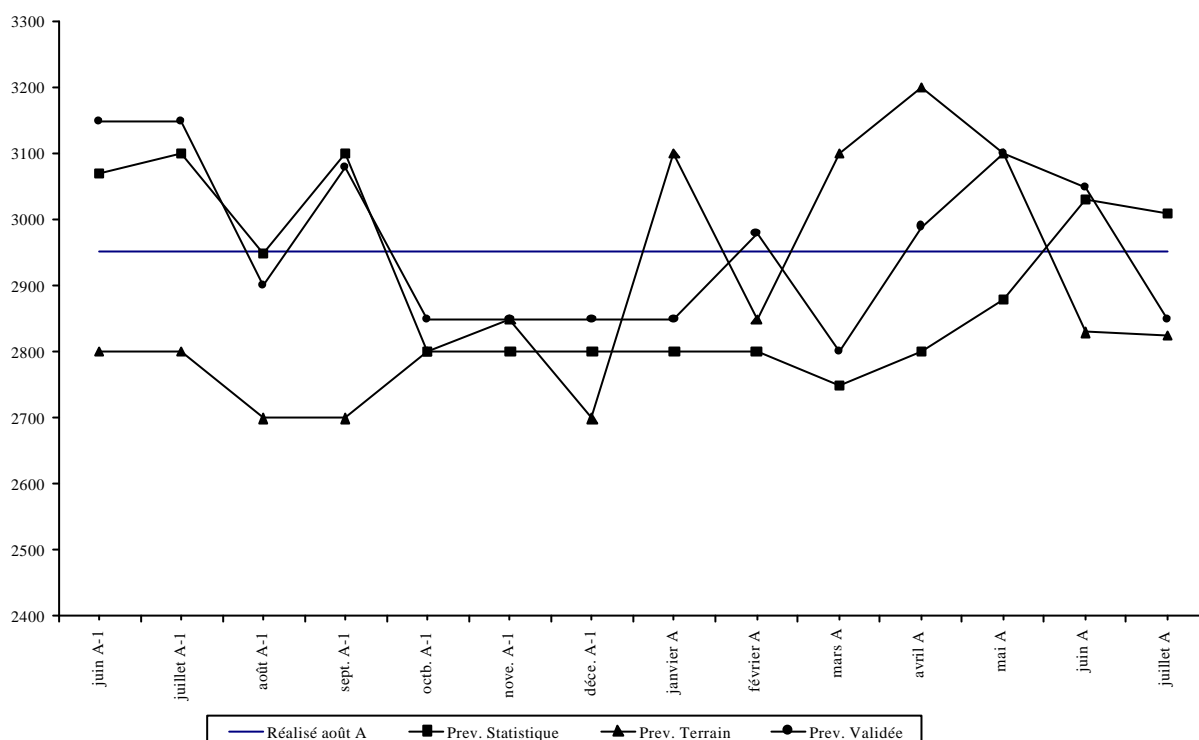
- les valeurs réalisées,
 - la prévision obtenue par différentes méthodes (prévision du terrain, statistique, etc.).
- Prenons l'exemple d'une entreprise qui gère trois sources différentes de prévision :

- statistique issue d'un modèle de prévision,
- terrain émanant de la consolidation des prévisions des commerciaux,
- validée en comité « prévision » (cf. chapitres 1 et 10).

Il peut être intéressant de comparer, sur un graphique, l'évolution dans le temps des différentes prévisions glissantes pour une même période de prévision.

Pour le mois d'août de l'année A, nous avons tracé l'évolution de trois prévisions glissantes différentes :

Graphique 23 – Comparaison entre réalisation et trois prévisions (statistique, terrain, validée)



Nous constatons que 12 mois avant l'échéance, la prévision « terrain » sous-estime le réalisé et reste relativement stable jusqu'à l'approche du mois d'août où elle subit des modifications brutales. La prévision statistique évolue jusqu'à convergence vers le réalisé. La prévision validée se situe le plus souvent entre la prévision statistique et terrain, ne sachant laquelle privilégier ...

3.3. Exemples d'outils d'alerte

Dans tout système de prévision portant sur un grand nombre de références, le gestionnaire doit pouvoir disposer d'outils d'alerte permettant de réagir par exception. Nous proposons ici une panoplie d'indicateurs opérationnels concernant un système mensuel de prévision de ventes.

La prévision historisée (les prévisions successives conservées) servant de référence est la prévision validée par l'utilisateur ; le nombre de prévisions historisées est égal à $T = 12$. Tous les indicateurs sont donc calculés sur les 12 dernières périodes de prévision. Les indicateurs présentés au début de ce chapitre ne font pas l'objet d'explications.

– Détection des valeurs anormales et filtrage

Si une observation de la série CVS en dehors de $\pm 2 s_x \Rightarrow$ suspicion de valeur anormale.

(s_x = écart-type de la série CVS).

– Amplitude de l'erreur de prévision anormalement élevée

Si $|EPS_t| / |MAD_t| > 3 \Rightarrow$ erreur de prévision anormalement élevée.

– Rupture de tendance

Indicateur *SUMEPS* (somme des erreurs cumulées, cf. chapitre 3)

Si $|SUMEPS_t| / |MAD_t| > 3 \Rightarrow$ risque de biais.

Ou test statistique

$$\text{Si } \frac{\sqrt{MSE_t}}{EPS} < \frac{\sqrt{T}}{2} \Rightarrow \text{risque de biais.}$$

– *Efficacité du modèle de prévision*

Indicateur *MSE (Mean Squared Error)*

Si $MSE_t > \text{variance des ventes CVS } (s_x^2) \Rightarrow$ modèle de prévision non efficace car la variance de l'erreur de prévision (l'aléa) est supérieure à la variance intrinsèque de la série. Le modèle de prévision n'a pas réduit l'aléa.

A chaque émission des prévisions le gestionnaire, par exception, va se consacrer en priorité aux articles dont l'une de ces anomalies a été constatée.