UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

E.507

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

RÉSEAU TÉLÉPHONIQUE ET RNIS

QUALITÉ DE SERVICE, GESTION DU RÉSEAU ET INGÉNIERIE DU TRAFFIC

MODÈLES DE PRÉVISION DU TRAFIC INTERNATIONAL

Recommandation UIT-T E.507

(Extrait du *Livre Bleu*)

NOTES

1 La Recommandation E.507 de l' UIT-T a été publiée dans le fascicule II.3 du Livre Bleu. Ce fichier est un extrait du Livre Bleu. La présentation peut en être légèrement différente, mais le contenu est identique à celui du Livre Bleu et les conditions en matière de droits d'auteur restent inchangées (voir plus loin).

2 Dans la présente Recommandation, le terme «Administration» désigne indifféremment une administration de télécommunication ou une exploitation reconnue.

Recommandation E.507[1)]

MODÈLES DE PRÉVISION DU TRAFIC INTERNATIONAL

1 Introduction

La mise au point de modèles économétriques et de modèles fondés sur des séries chronologiques et l'établissement de prévisions supposent que l'on connaisse bien les méthodes et les techniques pour pouvoir faire face à toute une série de situations différentes. Ainsi, la présente Recommandation expose certaines idées de base et renvoie le lecteur pour de plus amples détails aux publications de la liste des références. En tant que telle la présente Recommandation n'est pas censée être un guide complet pour la construction de modèles économétriques et de modèles fondés sur les séries chronologiques ainsi que pour l'établissement de prévisions.

La présente Recommandation donne également des indications pour construire divers modèles de prévision: identification du modèle, introduction de variables explicatives, ajustement pour corriger les irrégularités, estimation des paramètres, vérifications par diagnostic, etc.

La présente Recommandation décrit en outre diverses méthodes d'évaluation et de choix des modèles de prévision.

2 Construction du modèle de prévision

On peut aisément décomposer le processus de construction du modèle en quatre étapes successives. La première étape consiste à trouver une classe de modèles appropriée à la réalité. Le choix pourra porter par exemple sur des modèles simples, des modèles à lissage, des modèles à autorégression, des modèles à autorégression à moyenne glissante intégrée (ARIMA) ou encore des modèles économétriques. Avant de retenir une classe de modèles, il faut évaluer le rôle que peuvent jouer des variables externes. Si certaines variables externes particulières influent de manière appréciable sur le trafic demandé, elles doivent être retenues dans les modèles de prévision pourvu qu'une quantité suffisante de données historiques soit disponible.

L'étape suivante consiste à identifier un modèle provisoire dans la classe de modèles qui a été choisie. Si la classe retenue est trop vaste pour permettre un ajustement direct et sans inconvénient par rapport aux données, on peut utiliser des méthodes approximatives en vue d'identifier des sous-classes. De telles méthodes d'identification des modèles impliquent l'existence de données et une connaissance du système pour formuler avec parcimonie un choix approprié de sous-classe de modèles. Il peut également arriver que la procédure d'identification soit utilisée pour obtenir une estimation approximative préliminaire des paramètres du modèle. Ensuite, le modèle provisoire est ajusté aux données à l'aide d'une estimation des paramètres. En règle générale, on utilise des estimateurs par la méthode du maximum de vraisemblance ou par celle des moindres carrés.

L'étape suivante consiste à faire une vérification du modèle. Cette procédure, souvent appelée vérification par diagnostic, vise à déterminer le degré d'adéquation du modèle par rapport aux données et à indiquer des solutions possibles lorsque l'écart entre le modèle et les données est jugé trop important. Cette étape peut donc déboucher sur l'acceptation du modèle si l'adéquation obtenue est acceptable et, si elle ne l'est pas, c'est le signe que l'on peut désormais procéder à l'estimation et à la vérification par diagnostic d'autres modèles provisoires.

La figure 1/E.507 décrit les différentes étapes de la procédure de construction du modèle.

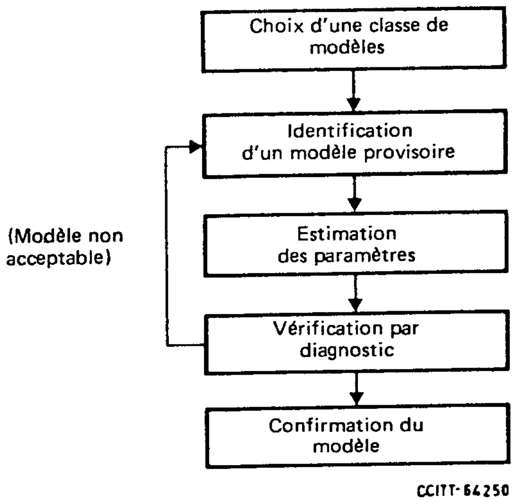

FIGURE 1/E.507
Différentes étapes du processus de construction d'un modèle

[1)] Le texte de l'ancienne Recommandation E.506 du *Livre rouge* a été réparti entre les nouvelles Recommandations E.506 et E.507 avec l'adjonction d'un grand nombre d'éléments nouveaux.

3 Divers modèles de prévision

L'objectif du présent § 3 est de donner un bref aperçu général des modèles de prévision les plus importants. Dans le Manuel du GAS 10 cité en [5] on trouvera une description plus détaillée des différents modèles.

3.1 *Modèles d'ajustement de la courbe*

Dans les modèles d'ajustement de la courbe, on extrapole la tendance du trafic en calculant les valeurs des paramètres d'une certaine fonction qui doit caractériser l'accroissement du trafic international dans le temps. Les calculs numériques de certains modèles d'ajustement de la courbe peuvent être effectués par la méthode des moindres carrés.

Voici des exemples des modèles d'ajustement de la courbe les plus couramment utilisés pour la prévision du trafic international.

Fonction linéaire: $$Y_t = a + bt \tag{3-1}$$

Fonction parabolique: $$Y_t = a + bt + ct^2 \tag{3-2}$$

Fonction exponentielle: $$Y_t = ae^{bt} \tag{3-3}$$

Fonction logistique: $$Y_t = \frac{M}{1 + ae^{bt}} \tag{3-4}$$

Fonction de Gompertz: $$Y_t = M(a)^{bt} \tag{3-5}$$

où

Y_t représente le trafic à l'instant t,

a, *b*, *c* sont des paramètres,

M est un paramètre représentant le niveau de saturation.

On trouvera aux figures 2/E.507 et 3/E.507 les différentes courbes de tendance.

Les courbes logistiques et les courbes de Gompertz diffèrent des courbes linéaires, paraboliques et exponentielles en ce qu'elles ont un niveau de saturation ou un niveau plafond. Pour de plus amples détails, voir [10].

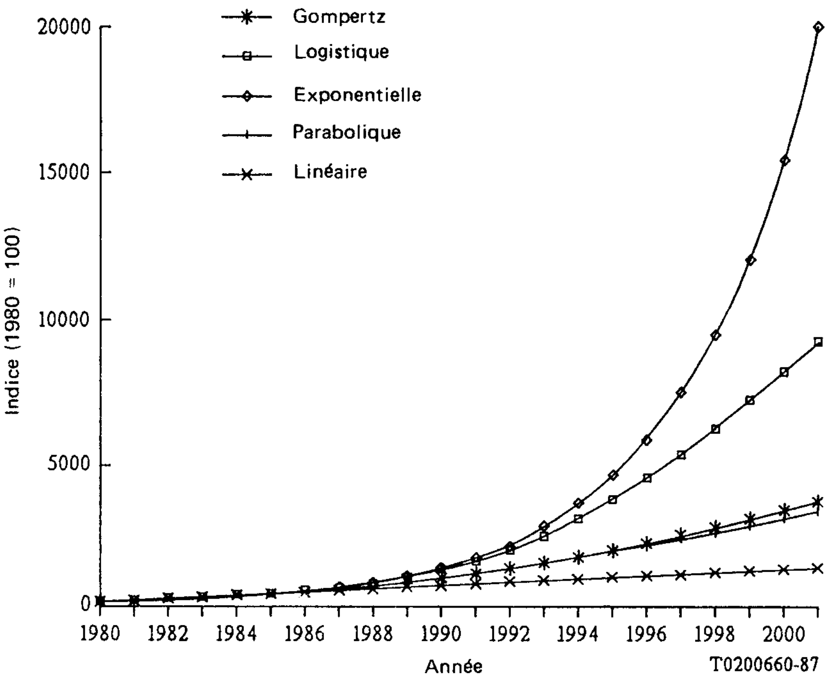

FIGURE 2/E.507

Exemple d'ajustement du trafic téléphonique international à l'aide de différents modèles

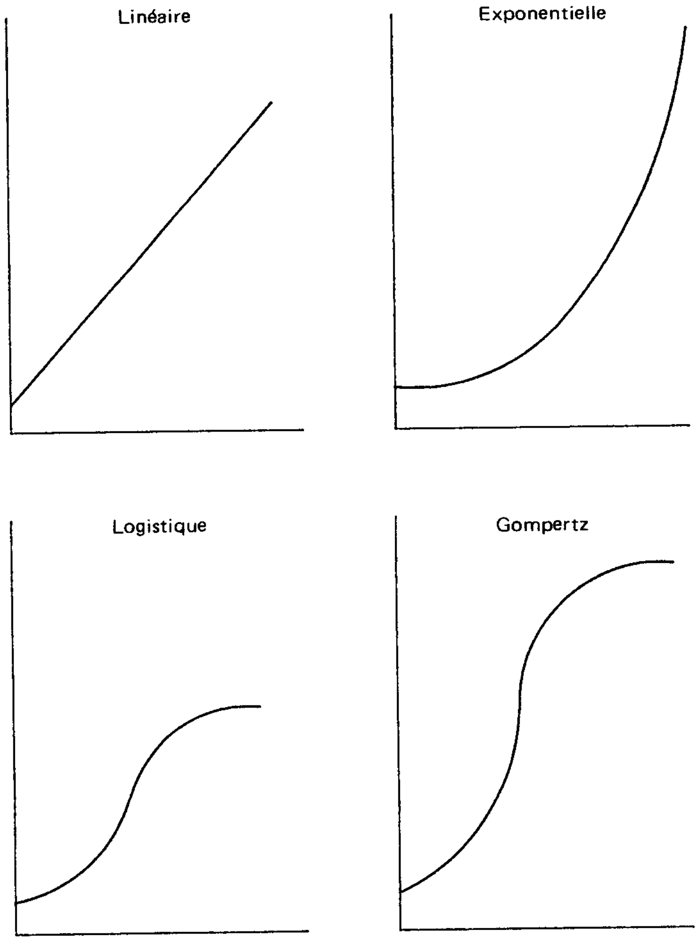

FIGURE 3/E.507

Exemples d'ajustement de courbes

3.2 *Modèles à lissage*

On peut, en appliquant un processus de lissage pour l'ajustement de la courbe, calculer les paramètres du modèle, et ajuster ainsi parfaitement les données actuelles, mais pas nécessairement les données obtenues longtemps auparavant.

Le procédé de lissage le plus connu est celui de la moyenne glissante. Le degré de lissage dépend du nombre d'observations les plus récentes qui sont prises en compte pour la moyenne et qui ont toutes le même poids.

Outre les modèles à moyenne glissante, il existe d'autres modèles à lissage fondés sur la pondération des observations. Les modèles les plus courants sont:

– les modèles à lissage exponentiel simples,
– les modèles à lissage exponentiel doubles,
– les modèles à régression décomptée,
– la méthode de Holt,
– les modèles saisonniers de Holt-Winters.

Par exemple, dans les modèles à lissage exponentiel, le poids donné aux observations antérieures décroît géométriquement avec l'ancienneté selon l'équation suivante:

$$\hat{\mu}_t = (1-a)Y_t + a\hat{\mu}_{t-1} \tag{3-6}$$

Y_t est le trafic mesuré à l'instant t,

$\hat{\mu}_t$ est le niveau prévu à l'instant t,

a est le facteur de décompte [et $(1 – a)$ est le paramètre de lissage].

L'impact des observations passées sur les prévisions dépend de l'importance du facteur de décompte.

L'emploi des modèles à lissage est tout indiqué dans les prévisions à court terme. Pour de plus amples renseignements, voir [1], [5] et [9].

3.3 *Modèles à autorégression*

Si l'on peut exprimer la demande de trafic X_t à l'instant t comme une combinaison linéaire d'anciennes observations équidistantes de la demande trafic antérieure, on est en présence d'un processus à autorégression. Le modèle est alors défini par l'expression:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + a_t \qquad (3\text{-}7)$$

où

a_t est le bruit blanc à l'instant t;

Φ_k, pour $k = 1, \ldots p$ sont les paramètres d'autorégression.

On utilise la notation $AR(p)$ pour le modèle puisqu'il est d'ordre p.

Une analyse à régression permet d'estimer les paramètres. Par suite de tendances communes, on constate en général une forte corrélation entre les variables exogènes (X_{t-1}, X_{t-2}, ... X_{t-p}) et, partant, les estimations des paramètres seront en corrélation. De plus, il est assez difficile de soumettre les estimations à des tests d'importance mathématique.

On peut encore calculer les coefficients d'autocorrélation empirique et utiliser ensuite les équations de Yule-Walker pour faire une estimation des paramètres [Φ_k]. Cette procédure peut être appliquée lorsque les séries chronologiques [X_t] sont stationnaires. Par contre, si les séries chronologiques ne sont pas stationnaires, on peut souvent les transformer en séries stationnaires, par exemple en les différenciant. La procédure d'estimation est décrite au § A.1.

3.4 *Modèles à autorégression à moyenne glissante intégrée (ARIMA)*

Les modèles à autorégression à moyenne glissante appelés modèles ARIMA, constituent une extension de la classe des modèles à autorégression, laquelle comprend également les modèles à moyenne glissante. Un modèle à moyenne glissante d'ordre q est défini par la formule:

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \ldots - \theta_q a_{t-q} \qquad (3\text{-}8)$$

où

a_t est le bruit blanc à l'instant t;

[θ_k] sont les paramètres de moyenne glissante.

En supposant que l'on peut exprimer le terme de bruit blanc utilisé dans les modèles à autorégression du § 3.3 avec un modèle à moyenne glissante, on obtient un modèle appelé modèle ARIMA (p, q) et défini comme suit:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \ldots - \theta_q a_{t-q} \qquad (3\text{-}9)$$

Le modèle ARIMA décrit une série chronologique stationnaire. Si la série chronologique n'est pas stationnaire, il faut distinguer les séries, comme suit:

Si Y_t est la série chronologique et B l'opérateur à décalage vers l'arrière on obtient:

$$X_t = (1 - B)^d \, Y_t \qquad (3\text{-}10)$$

où

d représente le nombre de différences pour obtenir le caractère stationnaire.

On obtient le nouveau modèle ARIMA (p, d, q) en insérant l'équation (3-10) dans l'équation (3-9).

La méthode utilisée pour analyser de telles séries chronologiques a été mise au point par MM. G. E. P. Box et G. M. Jenkins [3]. Pour analyser et prévoir lesdites séries chronologiques, il faut en général faire appel à un ensemble de programmes à séries chronologiques.

Comme le montre la figure 1/E.507, on procède à l'identification d'un modèle provisoire en déterminant les transformations nécessaires ainsi qu'un certain nombre de paramètres d'autorégression et de moyenne glissante. L'identification découle de la structure des autocorrélations et des autocorrélations partielles.

L'étape suivante, conformément à la figure 1/E.507, est celle de la procédure d'estimation par la méthode du maximum de vraisemblance. Malheureusement, les résultats sont difficiles à obtenir avec cette méthode parce qu'il faut résoudre un système d'équations non linéaire. Dans la pratique, les calculs de ce type doivent être effectués à l'aide d'un programme informatique. Le modèle de prévision est fondé sur l'équation (3-9) et le procédé consistant à faire des prévisions l unités de temps à l'avance est décrit au § A.2.

Les modèles de prévision décrits jusqu'ici sont des modèles à variable unique. On peut également faire intervenir des variables explicatives. Dans ce cas, le système sera décrit par un modèle à fonction de transfert. Les méthodes utilisées pour l'analyse des séries chronologiques dans un tel modèle sont assez identiques aux méthodes décrites plus haut.

On trouvera une description détaillée des modèles ARIMA en [1], [2], [3], [5], [11], [15] et [17].

3.5 *Modèles spatiaux d'état avec filtrage de Kalman*

Les modèles spatiaux d'état sont un moyen de représenter des processus discrets en temps à l'aide d'équations aux différences. Ce type de modélisation permet de convertir un modèle linéaire général quelconque en une forme appropriée pour l'estimation et la prévision récursives. On trouvera une description plus détaillée des modèles spatiaux d'état ARIMA en [1].

Pour un processus stochastique, la représentation peut prendre la forme suivante:

$$X_{t+1} = \Phi X_t + Z_t + \omega_t \tag{3-11}$$

et

$$Y_t = HX_t + v_t \tag{3-12}$$

où

X_t est un vecteur s de variables d'état au cours de la période t,

Z_t est un vecteur s d'événements déterministes,

Φ est une matrice de transition $s \times s$ pouvant en général dépendre de t,

ω_t est un vecteur s d'erreurs de modélisation aléatoires,

Y_t est un vecteur d de mesures au cours de la période t,

H est une matrice $d \times s$ appelée matrice d'observation, et

v_t est un vecteur d d'erreurs de mesure.

Aussi bien ω_t dans la formule (3-11) et que v_t dans la formule (3-12) sont des séquences aléatoires additives avec statistiques connues. La valeur prévue de chaque séquence est le vecteur zéro, et ω_t ainsi que v_t répondent aux conditions suivantes:

$$E\left[\omega_t \omega_j^T\right] = Q_t \delta_{tj} \quad \text{pour toutes les valeurs de } t, j,$$

$$\tag{3-13}$$

$$E\left[v_t v_j^T\right] = R_t \delta_{tj} \quad \text{pour toutes les valeurs de } t, j$$

où

Q_t et R_t sont des matrices définies non négatives,[2)]

et

δ_{tj} est le delta de Kronecker.

Q_t est la matrice de covariance des erreurs de modélisation et R_t est la matrice de covariance des erreurs de mesure; on suppose que les valeurs de ω_t et v_t ne sont pas corrélées et qu'elles sont assimilées à un bruit blanc. En d'autres termes:

$$E\left[v_t \omega_j^T\right] = 0 \text{ pour toutes les valeurs de } t, j, \tag{3-14}$$

et

$$E\left[v_t X_0^T\right] = 0 \text{ pour toutes les valeurs de } t. \tag{3-15}$$

Sur la base des hypothèses précédentes, on détermine $X_{t,t}$ de telle sorte que:

$$E\left[(X_{t,t} - X_t)^T (X_{t,t} - X_t)\right] = \text{minimum}, \tag{3-16}$$

où

$X_{t,t}$ est une estimation du vecteur d'état à l'instant t, et

X_t est le vecteur des vraies variables d'état.

La technique de filtrage de Kalman permet d'évaluer les variables d'état de manière récursive pour des applications en ligne. On procède comme suit: en supposant qu'il n'y a pas de variable explicative Z_t, chaque nouveau point de données devenant disponible est utilisé pour actualiser le modèle:

$$X_{t,t} = X_{t,t-1} + K_t (Y_t - HX_{t,t-1}) \tag{3-17}$$

où

K_t est la matrice de gain de Kalman, qui peut être calculée de manière récursive [18].

Intuitivement, la matrice de gain détermine le poids relatif qui sera donné à la dernière erreur de prévision observée pour compenser cette erreur. Afin d'établir une prévision d'intervalle d'anticipation k, on utilise la formule suivante:

$$X_{t+k,t} = \Phi^k X_{t,t} \tag{3-18}$$

dans laquelle

$X_{t+k,t}$ est une estimation de X_{t+k} observations données Y_1, Y_2..., Y_t.

---

2) A est une matrice définie non négative si, et seulement si, pour tous les vecteurs z, $z^T A z \geq 0$.

Les formules (3-17) et (3-18) montrent que la technique de filtrage de Kalman aboutit à une procédure de prévision commode de nature récursive et qu'elle permet d'obtenir des estimations sans distorsion de variance minimum pour le processus discret en temps qui est considéré.

On trouvera des détails supplémentaires dans [4], [5], [16], [18], [19] et [22].

Le filtrage de Kalman fonctionne bien lorsque les données examinées sont saisonnières. Les données relatives à la charge de trafic saisonnière peuvent être représentées par une série chronologique périodique. Ainsi, on peut obtenir un filtre de Kalman saisonnier en superposant un modèle de croissance linéaire et un modèle saisonnier.

On trouvera une description plus détaillée des techniques de filtrage de Kalman saisonnier dans [6] et [20].

3.6 *Modèles à régression*

Les formules (3-1) et (3-2) représentent des modèles à régression typiques où le trafic Y_t est la variable dépendante (ou explicative) et le temps *t* la variable indépendante.

Un modèle à régression décrit une relation linéaire entre la variable dépendante et la variable indépendante. Sur la base de certaines hypothèses, on peut utiliser la méthode des moindres carrés ordinaires pour évaluer les paramètres.

Un modèle comportant plusieurs variables indépendantes est appelé modèle à régression multiple. Il s'exprime ainsi:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t \qquad (3\text{-}19)$$

où

Y_t est le trafic à l'instant *t*,

β_i, $i = 0, 1, ..., k$ sont les paramètres,

X_{it}, $i_e = 1, 2, ..., k$ est la valeur des variables indépendantes à l'instant *t*

u_t est le terme d'erreur à l'instant *t*.

Les variables indépendantes ou explicatives susceptibles d'être utilisées dans le modèle à régression représentent par exemple la tarification, l'exportation et l'importation ou un degré d'automatisation. D'autres variables explicatives sont citées dans le § 2 (“Données de base pour les prévisions”) de la Recommandation E.506.

On trouvera une description détaillée des modèles à régression en [1], [5], [7], [15] et [23].

3.7 *Modèles économétriques*

Les modèles économétriques utilisent des équations établissant un rapport entre une variable devant être prévue (la variable dépendante ou endogène) et un certain nombre de variables socio-économiques (appelées variables indépendantes ou explicatives). La forme des équations doit refléter une relation de causalité prévue entre les variables. Selon une forme de modèle supposée, des données historiques ou des données de section transversale sont utilisées afin de prévoir les coefficients de l'équation. Si l'on suppose que le modèle conserve sa validité à terme, les estimations des valeurs futures des variables indépendantes peuvent être utilisées afin de prévoir les variables présentant un intérêt. Un exemple de modèle économétrique typique est donné dans l'annexe C.

On peut utiliser une large gamme de modèles et un certain nombre de méthodes pour estimer les coefficients (par exemple, les méthodes des moindres carrés, à paramètres variables, la régression non-linéaire, etc.). A bien des points de vues, la gamme des modèles économétriques disponibles est bien plus flexible que celle des autres modèles. Ainsi, pour ne prendre que quelques exemples, les modèles économétriques permettent d'incorporer les effets décalés, de pondérer les observations, d'intégrer les modèles résiduels ARIMA, de regrouper des informations sur différentes sections et de faire varier les paramètres.

L'un des principaux avantages de la construction d'un modèle de prévision économétrique est qu'il faut identifier correctement la structure ou le processus dont on extrait les données et déterminer des liens de causalité appropriés. Une identification explicite de la structure fait que les modèles économétriques se prêteront mieux que d'autres modèles à l'identification des erreurs commises dans la prévision.

Avec un modèle économétrique, il est facile d'isoler les modifications structurelles et d'éliminer les éléments non homogènes dans les données historiques ou d'en pondérer correctement l'influence. Par ailleurs, des modifications intéressant les facteurs qui influent sur les variables considérées peuvent être facilement intégrées dans les prévisions obtenues à partir d'un modèle économétrique.

On peut souvent élaborer un modèle économétrique relativement fiable avec moins d'observations qu'il n'en faut pour un modèle à séries chronologiques. Pour les modèles à régression par groupements, seules quelques observations intéressant plusieurs sections transversales suffisent si l'on veut construire un modèle applicable aux prévisions.

Cependant, il convient d'apporter beaucoup de soins à l'estimation d'un modèle afin de satisfaire aux hypothèses fondamentales des techniques décrites dans un certain nombre de documents énumérés dans la liste de références placée à la fin de la présente Recommandation. Par exemple, le nombre de variables indépendantes que l'on peut utiliser est limité par la quantité de données disponibles pour l'estimation du modèle. De même, il faut éviter les variables indépendantes entre lesquelles existent des corrélations. Il est parfois possible d'éviter la corrélation entre les variables par l'utilisation de données différenciées ou pour lesquelles il n'est pas tenu compte des tendances, ou encore par la transformation des variables. On trouvera des renseignements supplémentaires dans [8], [12], [13], [14] et [21].

4 Discontinuités dans l'accroissement du trafic

4.1 *Exemples de discontinuité*

Il est parfois difficile d'évaluer à l'avance l'ampleur d'une discontinuité. L'influence des facteurs qui causent des discontinuités est souvent étalée sur une période transitoire, et la discontinuité n'apparaît pas de façon évidente. Par ailleurs, il est difficile de déceler avec certitude les discontinuités dues à l'introduction, par exemple, de l'exploitation automatique dans le service international, toute modification du mode d'exploitation s'accompagnant en général d'autres changements (par exemple, des réductions de tarifs).

Le diagramme de la figure 4/E.507 permet de voir l'influence des discontinuités sur l'accroissement du trafic.

On connaît des cas de discontinuités qui ont doublé ou plus que doublé l'intensité du trafic. Il convient de relever aussi que la tendance d'accroissement du trafic peut changer après une discontinuité.

S'agissant de prévisions à court terme, il y aura peut-être intérêt à considérer la tendance du trafic entre les discontinuités; en revanche, pour des prévisions à long terme, il peut être préférable d'avoir recours à une estimation de tendance fondée sur des observations à long terme, compte tenu des discontinuités précédentes.

A part les fluctuations aléatoires dues à des pointes imprévisibles de trafic, à des dérangements, etc., les mesures de trafic sont sujettes à des fluctuations systématiques dues à des cycles journaliers ou hebdomadaires de l'intensité du trafic, à l'influence des différences d'heure, etc.

4.2 *Introduction de variables explicatives*

Le processus qui consiste à identifier des variables explicatives pour un modèle économétrique est probablement l'aspect le plus épineux de l'élaboration des modèles économétriques. Les variables explicatives utilisées dans un modèle économétrique mettent en évidence les principaux facteurs qui influent sur la variable considérée. On trouvera une liste de variables explicatives dans le § 2 de la Recommandation E.506.

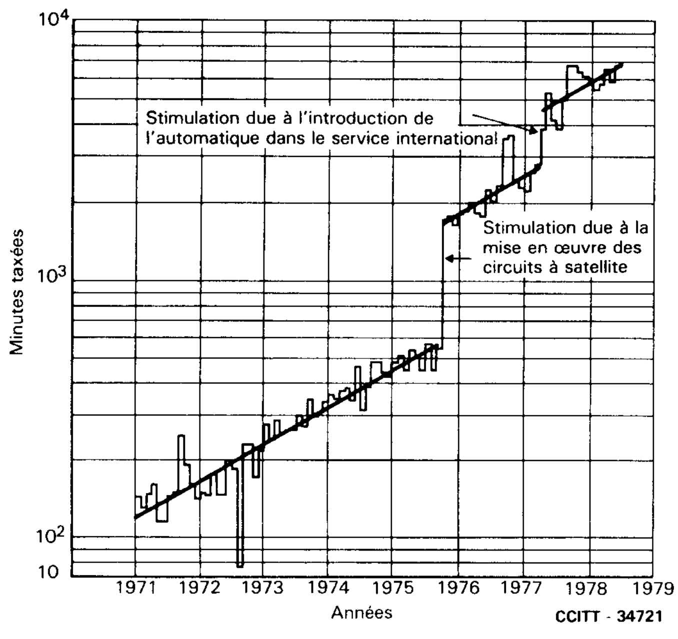

FIGURE 4 /E.507

Nombre de minutes taxées des communications téléphoniques au départ de l'Australie vers le Sri Lanka

La théorie économique est le point de départ de la sélection des variables et, plus précisément, la théorie de la demande fournit le cadre fondamental permettant d'élaborer le modèle général. Toutefois, la description de la structure ou du processus dont sont extraites les données conditionne souvent le choix des variables appelées à être retenues dans la série des variables explicatives. Par exemple, on peut être amené à incorporer des relations technologiques dans le modèle pour définir correctement la structure.

Le choix des variables explicatives est soumis à certains critères (par exemple, le $\overline{R}^2$, la statistique de Durbin-Watson, l'erreur quadratique moyenne, le niveau de prévision “ex-post”; pour obtenir les explications à ce sujet, on se reportera à la liste de références placée à la fin de la Recommandation). Malgré tout, des problèmes statistiques et/ou la disponibilité des données (historiques ou prévues) limitent l'éventail possible des variables explicatives et il faut souvent utiliser des variables supplétives. Par ailleurs, contrairement aux modèles statistiques purs, les modèles économétriques admettent des variables explicatives non seulement sur la base de critères statistiques mais aussi parce qu'il faut effectivement tenir compte de liens de causalité.

Un modèle économétrique entièrement spécifié pourra capter les points de variation. On n'observera aucune discontinuité dans la variable dépendante à moins que les paramètres du modèle subissent une transformation radicale en un laps de temps très bref. La présence de discontinuités dans l'accroissement du tarif téléphonique est une indication que le marché ou la structure technologique sous-jacents ont subi de profonds changements.

Pour capter des modifications durables de l'accroissement du trafic demandé, on peut utiliser la régression à paramètres variables ou encore introduire une variable dont la présence sert à expliquer la discontinuité (on peut, par exemple, introduire une variable relative à la publicité si l'on considère que la publicité est à l'origine de la modification structurelle). Le problème des discontinuités définitives ou graduelles ne peut pas être résolu par l'introduction de variables explicatives: dans ce cas, on utilisera des variables fictives.

4.3 *Introduction de variables fictives*

Avec un modèle économétrique, l'utilisation de variables qualitatives est souvent pertinente et, pour évaluer l'impact de ces variables, on introduit des variables fictives. Dans la technique des variables fictives, la valeur 1 correspond à la présence de l'attribut qualitatif qui influe sur la variable dépendante et la valeur 0 représente l'absence de cet attribut.

Les variables fictives sont donc appropriées lorsqu'on est en présence d'une discontinuité de la variable dépendante. Pour une variable fictive, on adoptera, par exemple, la valeur 0 pendant la période historique au cours de laquelle les appels ont été traités par une opératrice et la valeur 1 pour la période au cours de laquelle ils sont écoulés en service automatique.

On utilise souvent des variables fictives pour capter les points de variation de la variable dépendante qui sont imputables à des facteurs saisonniers ou lorsqu'il est nécessaire de supprimer l'effet qu'un élément non homogène produit sur les paramètres du modèle: il peut s'agir d'un accroissement soudain du trafic demandé à la suite d'une grève dans les services postaux ou d'un brusque repli imputable à des pannes d'installations quand les conditions météorologiques sont très mauvaises.

Il faut éviter d'utiliser les variables fictives sans discernement car:

1) elles tendent à absorber toute la puissance explicative pendant les discontinuités, et

2) elles aboutissent à une réduction du nombre de degrés de liberté.

5 Spécification des modèles

5.1 *Considérations générales*

Le présent § 5 décrit des méthodes qui permettent de tester l'importance mathématique des paramètres et de calculer des intervalles de confiance pour certains modèles de prévision mentionnés au § 3. Seront abordées en particulier les méthodes ayant trait aux analyses par régression et par séries chronologiques.

Tous les modèles de prévision économétriques présentés ici sont décrits comme des modèles à régression. Les modèles d'ajustement de courbe considérés au § 3.1 peuvent également être présentés comme des modèles à régression.

Le modèle exponentiel du type

$$Z_t = ae^{bt} \cdot u_t \tag{5-1}$$

peut être représenté sous une forme linéaire:

$$\ln Z_t = \ln a + bt + \ln u_t \tag{5-2}$$

ou encore

$$Y_t = \beta_0 + \beta_1 X_t + a_t \tag{5-3}$$

avec

$Y_t = \ln Z_t$

$\beta_0 = \ln a$

$\beta_1 = b$

$X_t = t$

$a_t = \ln u_t$ (bruit blanc).

5.2 *Autocorrélation*

Un bon modèle de prévision doit donner des résidus faiblement autocorrélés. Si la corrélation entre les résidus est importante, il se peut que l'estimation des paramètres et les prévisions soient médiocres. Pour vérifier si les erreurs sont corrélées, on détermine la fonction d'autocorrélation r_k, k = 1, 2, ... où r_k est l'autocorrélation estimée des résidus au décalage dans le temps k. Pour évaluer l'autocorrélation entre les résidus, une solution consiste à représenter la fonction d'autocorrélation et à procéder à un test de Durbin-Watson. La formule de la statistique de Durbin-Watson est:

$$\mathrm{D-W} = \frac{\sum_{t=2}^{N} (e_t - e_{t-1})^2}{\sum_{t=1}^{N} e_t^{\;2}} \tag{5-4}$$

où

e_t est le résidu estimé à l'instant t,

N est le nombre d'observations.

5.3 *Test d'importance mathématique des paramètres*

L'une des méthodes qui permettent d'évaluer un modèle de prévision consiste à analyser le rôle joué par différentes variables exogènes. Après avoir fait une estimation des paramètres dans le modèle à régression, il faut en contrôler l'importance mathématique.

Pour l'exemple de modèle économétrique décrit à l'annexe C, les valeurs estimées des paramètres sont indiquées. L'estimation de l'écart-type figure entre parenthèses sous ces valeurs. En règle générale, on considère que les paramètres sont significatifs si la valeur absolue des estimations est supérieure au double de la valeur de l'écart-type prévisionnel. Pour tester plus précisément l'importance mathématique des paramètres, on tiendra compte des lois de leurs estimateurs.

Le coefficient de corrélation multiple (ou coefficient de détermination) peut servir de critère pour l'ajustement de l'équation.

Le coefficient de corrélation multiple R^2 est donné par la formule:

$$R^2 = \frac{\sum_{i=1}^{N} (\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{N} (Y_i - \overline{Y})^2} \tag{5-5}$$

Si le coefficient de corrélation multiple est proche de l'unité, l'ajustement réalisé est satisfaisant, mais un coefficient R^2 élevé ne permet pas systématiquement d'obtenir des prévisions précises.

Avec une analyse par séries chronologiques, le modèle est abordé sous un angle différent. Comme nous l'avons indiqué au § 3.4, le nombre de paramètres autorégressifs de moyenne glissante dans un modèle ARIMA est obtenu par une procédure d'identification fondée sur la structure de la fonction d'autocorrélation et d'autocorrélation partielle.

L'estimation des paramètres et de leurs écarts-types est fondée sur une procédure d'estimation itérative non linéaire. Donc, avec un programme informatique d'analyse par séries chronologiques, il est possible de faire une estimation des paramètres en étudiant les écarts-types prévisionnels de la même manière que dans l'analyse par régression.

Pour effectuer un test global de l'ajustement, on utilise la statistique

$$Q_{N-d} = \sum_{i=1}^{N} r_i^2 \tag{5-6}$$

dans laquelle r_i est l'autocorrélation estimée au décalage dans le temps i, et d est le nombre de paramètres du modèle. Lorsque le modèle est approprié, Q_{N-d} est approximativement distribué selon la loi du χ^2 avec $N - d$ degrés de liberté. Pour contrôler l'adéquation, il est possible de comparer la valeur Q_{N-d} avec des fractiles de la loi du χ^2.

5.4 *Validité des variables exogènes*

Le principe des modèles de prévision économétriques repose sur l'utilisation d'une série de variables exogènes pour expliquer l'évolution de la variable endogène (c'est-à-dire du trafic demandé). Des prévisions relatives aux variables exogènes sont nécessaires si l'on veut prévoir le trafic demandé. Il est très important de noter qu'une variable exogène ne saurait être incorporée au modèle de prévision si le niveau de confiance obtenu pour la variable prévue est inférieur à celui de la prévision du trafic demandé.

Supposons que l'on connaît exactement le développement de la variable exogène, comme c'est par exemple le cas avec des modèles simples dans lesquels le temps est représenté par des variables explicatives. Si l'adéquation du modèle est bonne et si l'on obtient une distribution normale du bruit blanc avec une espérance mathématique nulle, il est possible de calculer les limites de confiance des prévisions, tâche facilement réalisable avec un programme d'ordinateur.

Par contre, il n'est généralement pas possible de prévoir avec exactitude les valeurs de la plupart des variables explicatives. La confiance de la prévision diminuera donc avec le nombre de périodes et, partant, les variables explicatives feront croître l'intervalle de confiance des prévisions proportionnellement au nombre de périodes de prévision. Dans un tel contexte, il est difficile de calculer un intervalle de confiance autour des valeurs prévues.

Si l'on peut exprimer le trafic demandé en utilisant un modèle à autorégression à moyenne glissante, aucune variable explicative n'est incluse dans le modèle. Donc s'il n'y a pas de variables explicatives dans le modèle, on peut calculer les limites de confiance des valeurs prévisionnelles en utilisant pour cela des programmes informatiques d'analyse par séries chronologiques.

5.5 *Intervalles de confiance*

Dans le contexte des prévisions, on appelle intervalle de confiance une construction statistique destinée à limiter le cadre des prévisions. Comme les modèles statistiques ne sont pas exempts d'erreurs, il y a une part d'incertitude dans l'estimation des paramètres. En d'autres termes, même si l'on a identifié le modèle de prévision approprié, des erreurs interviendront dans l'estimation des paramètres et dans les prévisions à cause des facteurs endogènes. Les intervalles de confiance tiennent compte de l'incertitude inhérente à l'estimation des paramètres.

Dans un modèle à liens de causalité, la prévision des variables explicatives est une autre source d'incertitude pour les prévisions concernant la série étudiée. Les intervalles de confiance sont alors impuissants et on néglige en général l'incertitude, bien qu'elle puisse être plus significative que l'incertitude propre à l'estimation des coefficients. Par ailleurs, l'incertitude imputable à d'éventuels chocs extérieurs n'apparaît pas dans les intervalles de confiance.

Pour un modèle à régression linéaire et statique, l'intervalle de confiance de la prévision dépend de la fiabilité des coefficients de régression, de la taille de la variance résiduelle et des valeurs des variables explicatives. L'intervalle de confiance de 95% correspondant à une valeur prévue Y_{N+1} est exprimé comme suit:

$$\hat{Y}_N(1) - 2\hat{\sigma} \le Y_{N+1} \le \hat{Y}_N(1) + 2\hat{\sigma} \tag{5-7}$$

où $\hat{Y}_N$ (1) est la prévision avec un intervalle d'anticipation et $\hat{\sigma}$ l'erreur type de la prévision.

On prévoit donc avec une probabilité de 95% que la valeur réelle de la série à l'instant $N + 1$ sera située à l'intérieur des limites de l'intervalle de confiance en supposant que la prévision des variables explicatives est totalement exempte d'erreurs.

6 Comparaison de différents modèles de prévision

6.1 *Contrôle de diagnostic – Evaluation des modèles*

Les essais et les contrôles de diagnostic sont des éléments importants dans la procédure d'élaboration des modèles. La qualité des modèles est déterminée par leurs résidus. Un bon modèle de prévision doit donner des résidus faiblement autocorrélés, la variance des résidus ne devant pas augmenter ou diminuer et l'espérance mathématique des résidus devant être nulle ou proche de zéro. La précision de la prévision dépend de la valeur des résidus, qui doivent être faibles.

En outre, les limites de confiance des estimations de paramètres et des prévisions doivent être relativement réduites. De même, pour qu'un modèle soit satisfaisant, l'erreur quadratique moyenne doit être faible par rapport aux résultats obtenus avec d'autres modèles.

6.2 *Prévision des niveaux et prévisions des variations*

Beaucoup de modèles économétriques sont élaborés à partir des niveaux des variables dépendantes et indépendantes. Comme les variables économiques évoluent simultanément dans le temps, on peut obtenir des coefficients de détermination élevés. La colinéarité constatée dans les niveaux des variables explicatives ne pose pas de problèmes lorsqu'un modèle est utilisé uniquement pour faire des prévisions car les caractéristiques de colinéarité obtenues à un moment donné se retrouvent par la suite. Toutefois, si l'on essaie d'évaluer les coefficients structurels (par exemple, l'élasticité de prix et l'élasticité de revenu), la colinéarité des variables explicatives, encore appelée multicolinéarité, fait que les coefficients prévisionnels donnent des résultats qui ne sont pas fiables.

Si l'on cherche à éviter le problème de la multicolinéarité et à obtenir des estimations de coefficient et des prévisions utilisables comme référence, il est possible de procéder à des changements de variables (première différence ou différence logarithmique équivalant à une modification en pourcentage) pour estimer un modèle et faire des prévisions à partir de ce modèle. Par des changements de variables à estimer, un modèle tend à supprimer les effets de mutlicolinéarité et à produire des estimations de coefficients plus fiables en reportant les effets communs des influences économiques sur les variables explicatives.

En réalisant des prévisions avec des niveaux et des changements de variables explicatives, on peut éventuellement obtenir de meilleurs résultats avec un processus de réconciliation, c'est-à-dire en ajustant les modèles pour que les deux séries de prévisions donnent des résultats équivalents.

6.3 *Prévisions "ex-post"*

La méthode de prévision "ex-post" consiste à utiliser un modèle estimé sur la base d'un sous-échantillon des données qui commencent avec la première observation et prennent fin plusieurs périodes avant la dernière observation. Pour faire des prévisions "ex-post", on utilise les valeurs réelles des variables explicatives et, si l'on prend les valeurs prévues de ces variables, il est possible de mesurer l'erreur introduite quand les prévisions des variables explicatives sont incorrectes.

Les prévisions "ex-post" servent à évaluer le niveau d'exactitude du modèle; on compare pour cela les valeurs prévues avec les valeurs réelles de la période allant de la fin du sous-échantillon à la dernière observation. Les prévisions "ex-post" permettent d'évaluer le niveau d'exactitude en ce qui concerne:

1) les écarts de pourcentage entre les valeurs prévues et réelles,

2) les points critiques dans le comportement du modèle,

3) le comportement systématique des écarts.

Les écarts entre les valeurs prévues et réelles permettent de porter un jugement global sur la précision du modèle. Les déplacements systématiques dans les écarts peuvent fournir des renseignements qui permettront de respécifier le modèle ou d'ajuster les prévisions compte tenu du déplacement observé. Pour évaluer l'exactitude du modèle, il est tout aussi important de connaître l'existence de points critiques dans le comportement du modèle: cela permet de savoir dans quelle mesure le modèle est capable de prévoir les changements avec le mouvement de la variable dépendante. D'autres critères qui permettent d'évaluer l'exactitude des prévisions sont abordés plus loin.

6.4 *Critères d'évaluation de la qualité des prévisions*

Même si le modèle concorde bien avec les données historiques, la concordance est parfois moins bonne quand on compare les prévisions et les données futures qui ne sont pas utilisées pour évaluer les paramètres. Une comparaison des prévisions et des valeurs observées peut donc fournir des renseignements supplémentaires sur la qualité du modèle. Supposons la série chronologique $Y_1, Y_2, ..., Y_N, Y_{N+1}, ..., Y_{N+M}$.

Les M dernières observations sont éliminées de la série chronologique et de la procédure d'élaboration du modèle. L'erreur de prévision avec un intervalle d'anticipation est donnée par la formule:

$$e_{N+t} = \hat{Y}_{N+t} - \hat{Y}_{N+t-1}(1) \qquad t = 1, 2, ..., M \tag{6-1}$$

où

$\hat{Y}_{N+t-1}(1)$ est la prévision avec un intervalle d'anticipation.

Erreur moyenne

L'erreur moyenne, EM, est définie par

$$\text{EM} = \frac{1}{M}\sum_{t=1}^{M} e_{N+t} \tag{6-2}$$

L'erreur moyenne est un critère permettant d'évaluer la déviation de prévision. Etant donné que l'espérance mathématique des résidus doit être nulle, un écart important par rapport à zéro dénote la présence de déviations dans les prévisions.

Erreur relative moyenne

L'erreur relative moyenne, ERM, est définie par

$$\text{ERM} = \frac{100}{M}\sum_{t=1}^{M} \frac{e_{n+t}}{Y_{N+t}} \tag{6-3}$$

En outre, cette statistique met en évidence les déviations possibles de prévision. Le critère de l'erreur relative moyenne permet de mesurer l'écart relatif des déviations. Il n'est pas recommandé d'utiliser ce critère quand les valeurs observées sont faibles.

Erreur quadratique moyenne

L'erreur quadratique moyenne, EQM, de la prévision est définie par:

$$\text{EQM} = \left[\frac{1}{M}\sum_{t=1}^{M} e_{N+t}^2\right]^{1/2} \tag{6-4}$$

L'erreur quadratique moyenne est le critère le plus communément utilisé pour évaluer la précision des prévisions.

Erreur moyenne absolue

L'erreur moyenne absolue, EMA, est définie par:

$$\text{EMA} = \frac{1}{M}\sum_{t=1}^{M} \left|e_{N+t}\right| \tag{6-5}$$

Coefficient d'inégalité de Theil

Le coefficient d'inégalité de Theil est défini par:

$$U = \left[\sum_{t=1}^{M} \frac{e_{N+t}^2}{Y_{N+t}^2} \right]^{1/2} \tag{6-6}$$

On préfère utiliser le coefficient de Theil U pour mesurer l'exactitude des prévisions parce que l'erreur entre les valeurs prévues et réelles peut être décomposée en différentes erreurs, dont l'origine est:

1) la tendance centrale,

2) la variation inégale entre les modifications prévues et réelles, et

3) la covariation incomplète des modifications prévues et réelles.

On peut utiliser cette décomposition pour ajuster le modèle afin d'en améliorer la précision.

Un modèle de prévision doit posséder une autre qualité: pouvoir capter les points critiques. En d'autres termes, une prévision doit pouvoir changer de direction pendant la période au cours de laquelle la série réelle qui est étudiée change de direction elle aussi. Si l'on estime un modèle sur une longue période comportant plusieurs points critiques, une analyse prévisionnelle “ex-post” peut généralement mettre en évidence l'incapacité du modèle à capter avec exactitude les séries réelles qui présentent des points critiques.

7 Choix du modèle de prévision

7.1 *Qualité des prévisions*

Même si l'on choisit généralement un modèle de prévision en fonction de ses caractéristiques de prévision, il faut être attentif à d'autres facteurs. Pour un modèle économétrique, il faut notamment envisager la longueur de la période de prévision, la forme fonctionnelle et l'exactitude des prévisions de variables explicatives.

La longueur de la période de prévision influe sur le choix d'un modèle par rapport à un autre, au même titre que les limitations relatives aux données historiques et le but du modèle de prévision. Par exemple, les modèles à autorégression à moyenne glissante intégrée (ARIMA) conviendront peut-être pour les prévisions à court terme si l'on dispose d'une quantité suffisante de données historiques et si la stabilité ou la causalité n'ont pas d'importance. Par ailleurs, si la structure dont on extrait les données est difficilement identifiable, il ne reste plus qu'à utiliser un modèle de prévision fondé sur des données historiques concernant la variable considérée.

Il faut également tenir compte de la forme fonctionnelle d'un modèle de prévision. De même qu'un modèle plus complexe peut réduire l'erreur de spécification du modèle, il augmentera généralement les effets des erreurs de données d'une façon considérable. La forme du modèle doit être choisie de façon à composer un compromis entre ces sources d'erreurs.

En ce qui concerne les variables explicatives, la disponibilité des prévisions et le niveau de fiabilité sont encore des facteurs qui influent sur le choix du modèle de prévision. Un modèle supérieur fondé sur des variables explicatives qu'il est impossible de prévoir avec précision peut se révéler inférieur à un modèle moyen dont il est possible de prévoir les variables explicatives avec exactitude.

Quand la stabilité du marché joue un rôle pertinent, il convient d'utiliser des modèles économétriques capables de capter les changements structurels. Si l'on doit tenir compte de la causalité, il est impossible d'utiliser un modèle simple ou un modèle ARIMA comme outil de prévision. Ces mêmes modèles seront également inutilisables en présence d'une quantité insuffisante de données historiques. Enfin, quand un modèle est conçu pour prévoir les conséquences de changements intéressant les facteurs qui influent sur la variable considérée, les modèles à série chronologiques peuvent ne pas convenir (à l'exception bien sûr des modèles à fonction de transfert et à séries chronologiques multiples).

7.2 *Durée de la période de prévision*

Dans le cas d'extensions normales (équipements de commutation, circuits), une période de prévision d'environ six ans est nécessaire. En revanche, il peut être nécessaire d'envisager des périodes de prévision plus longues lorsqu'il s'agit des plans pour de nouveaux câbles ou d'autres supports de transmission, de même que pour de grandes installations. Les estimations à long terme sont nécessairement moins précises que les estimations à court terme, mais sont néanmoins acceptables.

En ce qui concerne les prévisions faites avec un modèle statistique, la durée de la période de prévision dépend entièrement:

a) des données historiques disponibles,

b) de l'usage auquel est destinée la prévision,

c) de la structure du marché dont sont extraites les données,

d) du modèle de prévision utilisé, et

e) de la fréquence des données.

Les données historiques disponibles dépendent de la période au cours de laquelle elles ont été rassemblées et de la fréquence de ce regroupement (ou de la durée de la période au cours de laquelle des données sont rassemblées). Une petite base de données de référence au passé ne peut permettre des prévisions que sur un court intervalle. Par exemple, avec 10 ou 20 observations on peut utiliser un modèle pour prévoir 4 à 5 périodes après l'échantillon (c'est-à-dire, vers le futur). Mais avec 150 à 200 observations, des prévisions potentiellement fiables peuvent être obtenues sur 30 à 50 périodes après l'échantillon, le reste ne change pas.

Il ne fait aucun doute que l'objectif de la prévision influe sur le nombre des périodes couvertes par cette prévision. Si l'on veut planifier des installations à long terme, il faut effectuer des prévisions portant sur 15 à 20 ans ou plus à partir de l'échantillon. En ce qui concerne l'évolution des taxes téléphoniques, des prévisions portant sur 2 ou 3 années seulement peuvent peut-être suffire. Pour les réaménagements de voies d'acheminement, il pourrait suffire de prendre des périodes de quelques mois à partir de l'échantillon.

La stabilité – ou l'instabilité – du marché affecte également la durée de la période de prévision. Avec un marché dont la structure est stable, on pourrait concevoir d'étaler la période de prévision sur une durée égale à la période historique, mais dans le cas contraire la situation n'est pas aussi avantageuse car la période de prévision ne peut alors englober que quelques périodes vers le futur.

La nature d'un modèle de prévision fait qu'elle affecte la décision sur les limites pouvant raisonnablement être choisies dans le futur pour la période de prévision. Quand il s'agit de prévisions à long terme, les modèles structuraux tendent à fournir de meilleurs résultats que les autres modèles alors que n'importe quel modèle semble convenir de la même manière pour les prévisions à court terme.

Il convient de noter que, même si le but de la prévision et le modèle utilisé influent sur la durée de prévision, le nombre de périodes à couvrir joue un rôle crucial dans le choix du modèle de prévision et pour l'utilisation des résultats.

ANNEXE A

(à la Recommandation E.507)

Description des procédures de prévision

A.1 *Estimation des paramètres autorégressifs*

L'autocorrélation empirique au décalage dans le temps *k* est exprimée comme suit:

$$r_k = \frac{v_k}{v_0} \tag{A-1}$$

où

$$v_k = \frac{1}{N-1}\sum_{t=1}^{N-k}(X_t - \overline{X})(X_{t+k} - \overline{X}) \tag{A-2}$$

et

$$\overline{X} = \frac{1}{N}\sum_{t=1}^{N} X_t \qquad \text{(A-3)}$$

N étant le nombre total d'observations effectuées.

La relation entre $[r_k]$ et les estimations $[\hat{\Phi}_k]$ de $[\Phi_k]$ est obtenue à l'aide des équations de Yule-Walker:

$$\begin{aligned} r_1 &= \hat{\Phi}_1 + \hat{\Phi}_2 r_1 + \ldots + \hat{\Phi}_p r_{p-1} \\ r_2 &= \hat{\Phi}_1 r_1 + \hat{\Phi}_2 r_2 + \ldots + \hat{\Phi}_p r_{p-2} \\ &. \\ &. \\ &. \\ r_p &= \hat{\Phi}_1 r_{p-1} + \hat{\Phi}_2 r_{p-2} + \ldots + \hat{\Phi}_p \end{aligned} \qquad \text{(A-4)}$$

Donc, les estimateurs $[\Phi_k]$ peuvent être obtenus par résolution de ce système d'équations.

Aux fins des calculs, la procédure récursive décrite dans la suite du texte peut remplacer la procédure qui consiste à résoudre directement les équations. Supposons que $[\hat{\Phi}_{k,j}]_j$ sont les estimateurs des paramètres au décalage dans le temps j = 1, 2, ..., k en admettant que k est le nombre total des paramètres. On calcule alors les estimateurs $[\hat{\Phi}_{k+1,j}]_j$ de la manière suivante:

$$\hat{\Phi}_{k+1,k+1} = \frac{r_{k+1} \sum_{j=1}^{k} \Phi_{k,j} r_{k+1-j}}{1 - \sum_{j=1}^{k} \hat{\Phi}_{k,j} r_j} \qquad \text{(A-5)}$$

$$\hat{\Phi}_{k+1,j} = \hat{\Phi}_{k,j} - \hat{\Phi}_{k+1,k+1} \hat{\Phi}_{k,k-j+1} \qquad j = 1,2,\ldots,k \qquad \text{(A-6)}$$

A partir de la définition $\hat{\Phi}_{p,j} = \hat{\Phi}_j$, j = 1, 2, ..., p, on exprime la prévision du trafic demandé à l'instant t+1 en écrivant:

$$X_{t+1} = \hat{\Phi}_1 X_t + \hat{\Phi}_2 X_{t-1} + \ldots + \hat{\Phi}_p X_{t-p} \qquad \text{(A-7)}$$

A.2 *Prévisions faites avec des modèles ARIMA*

La prévision effectuée l unités de temps à l'avance s'obtient de la manière suivante:

$$\begin{aligned} \hat{X}_t(l) = {} & \hat{\Phi}_1 [X_{t+l-1}] + \hat{\Phi}_2 [X_{t+l-2}] \\ & + \ldots + \hat{\Phi}_p [X_{t+l-p}] \\ & + [a_{t+l}] - \hat{\theta}_1 [a_{t+l-1}] \\ & - \hat{\theta}_2 [a_{t+l-2}] - \ldots - \hat{\theta}_q [a_{t+l-q}] \ , \end{aligned} \qquad \text{(A-8)}$$

où $$[\hat{X}_j] = \begin{cases} \hat{X}_t(j-t) & \text{si } j > t \\ X_j & \text{si } j \le t \end{cases} \quad \text{(A-9)}$$

$$[a_j] = \begin{cases} 0 & \text{si } j > t \\ X_j - \hat{X}_j & \text{si } j \le t, \end{cases} \quad \text{(A-10)}$$

cela signifie que $[X_j]$ est défini comme une prévision lorsque $j > t$, ou sinon comme une observation réelle, et que $[a_j]$ est de valeur nulle lorsque $j > t$ puisque le bruit blanc a une espérance mathématique nulle. Si les observations sont connues ($j \le t$), alors $[a_j]$ correspond à la valeur résiduelle.

ANNEXE B

(à la Recommandation E.507)

Filtre de Kalman pour un modèle à tendance linéaire

Pour modéliser le trafic téléphonique, on suppose que la structure de la demande ne subit aucune modification déterministe. On peut alors procéder à la modélisation en donnant la valeur zéro à la composante déterministe Z_t. Le modèle spatial d'état prend alors la forme générale suivante:

$$\begin{aligned} X_{t+1} &= \varphi X_t + \omega_t \\ Y_t &= H X_t + v_t \end{aligned} \quad \text{(B-1)}$$

où

X_t est un vecteur s de variables d'état au cours de la période t,

Y_t est un vecteur de mesures au cours de l'année t,

φ_t est une matrice de transition s x s pouvant en général dépendre de t,

et

ω_t est un vecteur s d'erreurs de modélisation aléatoires,

v_t est l'erreur de mesure au cours de l'année t.

Pour modéliser la demande de trafic téléphonique, on adapte un modèle simple de variables à deux états et à donnée unique défini par:

$$X_{t+1} = \begin{bmatrix} x_{t+1} \\ \dot{x}_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} + \begin{bmatrix} \omega_t \\ \dot{\omega}_t \end{bmatrix} \quad \text{(B-2)}$$

et

$$y_t = x_t + v_t \quad \text{(B-3)}$$

où

x_t est la charge réelle au cours de l'année t,

$\dot{x}_t$ est la croissance incrémentielle réelle au cours de l'année t,

y_t est la charge mesurée au cours de l'année t,

v_t est l'erreur de mesure au cours de l'année t.

Ainsi, dans notre modèle

$$\varphi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad \text{et } H = 1. \tag{B-4}$$

La prévision avec un intervalle d'anticipation s'exprime comme suit:

$$X_{t+1,t} = \begin{bmatrix} x_{t+1,t} \\ \dot{x}_{t+1,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t,t} \\ \dot{x}_{t,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t,t-1} + \alpha_t (y_t - x_{t,t-1}) \\ \dot{x}_{t,t-1} + \beta_t (y_t - x_{t,t-1}) \end{bmatrix} \tag{B-5}$$

Dans cette formule,

$X_{t+1,t}$ est la prévision des variables d'état au cours de la période $t + 1$ compte tenu des observations effectuées pendant l'année t.

Les coefficients α_t et β_t sont les matrices de gain de Kalman au cours de l'année t. En réécrivant les formules ci-dessus, on obtient:

$$x_{t,t} = (1 - \alpha_t) x_{t,t-1} + \alpha_t y_t \tag{B-6}$$

et

$$\dot{x}_{t,t} = (1 - \beta_t)\, \dot{x}_{t,t-1} + \beta_t (y_t - x_{t-1,t-1}) \tag{B-7}$$

Le filtre de Kalman crée une tendance linéaire pour chaque série chronologique faisant l'objet de prévisions sur la base de l'observation effectuée ou des mesures de la demande de trafic et de la prévision de l'année précédente pour cette demande. On combine le résultat de la charge de trafic observée et prévue pour obtenir à la fois une charge lissée qui correspond au niveau du processus et un incrément de croissance lissé. Les valeurs de gain de Kalman α_t et β_t peuvent être fixes ou adaptables. Dans [16], Moreland présente une méthode pour choisir des paramètres fixes et sûrs garantissant des résultats suffisants quels que soient le bruit du système, les erreurs de mesure et les conditions initiales. On trouvera des renseignements additionnels sur les conditions appropriées du choix de ces paramètres dans [6], [20] et [22].

ANNEXE C

(à la Recommandation E.507)

Exemple de modèle économétrique

Pour illustrer le fonctionnement d'un modèle économétrique, nous avons choisi l'exemple du nombre de minutes taxées pour le trafic acheminé vers le Brésil à partir des États-Unis. Trois raisons expliquent notre choix; nous voulions en effet:

a) donner un exemple d'introduction des variables explicatives;

b) mettre en évidence les difficultés qui surgissent lorsqu'un modèle est utilisé à la fois pour estimer la structure et pour faire des prévisions, et

c) montrer comment des transformations peuvent affecter les résultats.

La demande en minutes taxées pour le trafic acheminé vers le Brésil à partir des États-Unis (*MIN*) est estimée au moyen d'une équation logarithmique linéaire qui inclut des messages échangés dans le sens États-Unis vers Brésil (*MSG*), un indice des prix téléphoniques en termes réels (*RPI*), le revenu des particuliers aux États-Unis aux prix de 1972 (*YP*72) et les échanges commerciaux bilatéraux en termes réels entre les États-Unis et le Brésil (*RTR*) comme variable explicative. Le modèle est représenté par la formule:

$$\ln(MIN)_t = \beta_0 + \beta_1 \ln(MSG)_t + \beta_2 \ln(RPI)_t + \beta_3 \ln(YP72)_t + \beta_4 \ln(RTR)_t + u_t \qquad \text{(C-1)}$$

où u_t est le terme d'erreur de la régression et dans laquelle on s'attend à obtenir les inégalités suivantes: $\beta_1 > 0$, $\beta_2 < 0$, $\beta_3 > 0$ et $\beta_4 > 0$.

En utilisant une régression non linéaire (régression de la crête) lorsque les problèmes de multicolinéarité sont importants, nous estimons l'équation sur l'intervalle compris entre 1971/1 (c'est-à-dire le 1er trimestre de 1971) et 1979/4, ce qui donne le résultat suivant:

$$\ln(MIN)_t = -3{,}489 + \underset{(0{,}035)}{0{,}619} \ln(MSG)_t - \underset{(0{,}095)}{0{,}447} \ln(RPI)_t + \underset{(0{,}269)}{1{,}166} \ln(YP72)_t + \underset{(0{,}084)}{0{,}281} \ln(RTR)_t \qquad \text{(C-2)}$$

$$\overline{R}^2 = 0{,}985, \quad SER = 0{,}083, \quad D\text{-}W = 0{,}922, \quad k = 0{,}10 \qquad \text{(C-3)}$$

où $\overline{R}^2$ est le coefficient de détermination ajusté, *SER* est l'erreur type de la régression, *D-W* est la statistique de Durbin-Watson et *k* est la constante de régression non linéaire. Les valeurs entre parenthèses placées sous l'équation sont les écarts-types des valeurs estimées des paramètres $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$.

Dans le cas présent, il a fallu introduire les messages à titre de variable explicative car la qualité de transmission s'est améliorée depuis le milieu des années 70 et les taxes téléphoniques ont augmenté tandis que, parallèlement, la forte croissance enregistrée sur le marché considéré commençait à s'estomper. Par ailleurs, la modification des taxes dans chacun des deux pays ou encore le revenu des particuliers aux États-Unis en termes réels n'auraient pas pu expliquer les taux de croissance enregistrés sur certaines périodes. Le comportement des messages dans l'équation logarithmique linéaire des minutes a permis d'expliquer tous ces facteurs.

Comme le modèle joue un double rôle (estimation de la structure et prévisions), on introduit au moins une variable de plus que si le modèle devait être utilisé seulement pour les prévisions. L'introduction de variables explicatives supplémentaires pose de sérieux problèmes de multicolinéarité et nécessite l'emploi d'une régression non linéaire afin de réduire le coefficient $\overline{R}^2$ et la valeur de la statistique de Durbin-Watson, ce qui a pour effet de diminuer quelque peu la puissance de prévision du modèle.

Dans le cas du trafic acheminé vers le Brésil à partir des Etats-Unis, l'analyse prévisionnelle “ex-post” montre quelles sont les conséquences de la transformation des variables d'un modèle. Les écarts obtenus avec les niveaux des variables sont plus importants que ceux des logarithmes des variables qui ont été utilisés pour obtenir une meilleure adéquation (l'erreur quadratique moyenne estimée pour le modèle logarithmique linéaire à régression est de 0,119 827). Les résultats des prévisions figurent dans le tableau C-1/E.507 sous forme de niveaux et de logarithmes.

TABLEAU C-1/E.507

	Logarithmes			Niveaux		
	Valeur Prévue	Valeur réelle	Ecart en (%)	Niveaux	Valeur prévue	Ecart en (%)
1980: 1	14,858	14,938	–0,540	2 836 269	3 073 697	–7,725
2	14,842	14,972	–0,872	2 791 250	3 180 334	–12,234
3	14,916	15,111	–1,296	3 005 637	3 654 092	–17,746
4	14,959	15,077	–0,778	3 137 698	3 529 016	–11,089
1981: 1	15,022	15,102	–0,535	3 341 733	3 621 735	–7,731
2	14,971	15,141	–1,123	3 175 577	3 762 592	–15,601
3	15,395	15,261	0,879	4 852 478	4 244 178	14,333
4	15,405	15,302	0,674	4 901 246	4 421 755	10,844
1982: 1	15,365	15,348	0,110	4 709 065	4 630 238	1,702
2	15,326	15,386	–0,387	4 528 947	4 807 901	–5,802

Références

[1] ABRAHAM (A.) et LEDOLTER (J.): Statistical methods for forecasting, *J. Wiley*, New York, 1983.

[2] ANDERSON (O. D.): Time series analysis and forecasting. The Box-Jenkins approach. *Butterworth*, London, 1976.

[3] BOX (G. E. P.) et JENKINS (G. M.): Time Series Analysis: Forecasting and Control, *Holden-Day*, San Francisco, 1976.

[4] BROWN (R. G.): Introduction to random signal analysis and Kalman Filtering. *John Wiley & Sons*, New York, 1983.

[5] Manuel du CCITT: *Données de planification et méthodes de prévision*, Vol. I et II, UIT, Genève, 1988.

[6] CHEMOUIL (P.) et GARNIER (B.): An Adaptive Short-Term Traffic Forecasting Procedure Using Kalman Filtering. *ITC 11*, Tokyo, 1985.

[7] DRAPER (N.) et SMITH (H.): Applied Regression Analysis, Second Edition, *John Wiley & Sons*, New York, 1981.

[8] DUTTA (M.): Econometric Methods, *South-Western Publishing Co.*, Cincinnati, 1975.

[9] GARDNER (E. S. Jr.): Exponential smoothing the state of art. *Journal of forecasting*, 4, pp. 1 à 28, 1985.

[10] GILCHRIST (W.): Statistical forecasting. *John Wiley & Sons*, New York, 1976.

[11] GRANGER (C. W. J.) et NEWBOLD (P.): Forecasting Economic Time Series, *Academic Press*, New York, 1977.

[12] JOHNSTON (J.): Econometric Methods, Second Edition, *McGraw-Hill*, New York, 1972.

[13] JUDGE (G. G.) *et al.*: The Theory and Practice of Econometrics, *John Wiley & Sons*, New York, 1980.

[14] KMENTA (J.): Elements of Econometrics, *Macmillan Publishing Company*, New York, 1971.

[15] MAKRIDAKIS (S.), WHEELWRIGHT (S.C.), McGEE (V.E.): Forecasting Methods and applications Second Edition. *John Wiley & Sons*, New York, 1983.

[16] MORELAND (J.P.): A robust sequential projection algorithm for traffic load forecasting. *The Bell Technical Journal*, vol. 61, n° 1, 1982.

[17] NELSON (C. R.): Applied Time Series Analysis for Managerial Forecasting, *Holden-Day*, San Francisco, 1973.

[18] PACK (C. D.) et WHITAKER (B. A.): Kalman Filter models for network forecasting. *The Bell Technical Journal*, Vol. 61, No. 1, pp. 1 à 9, 1982.

[19] SORENSON (H. W.): Kalman Filtering techniques. Advances in control systems theory and applications. *Academic Press*, Vol. 3, pp. 219 à 292, 1966.

[20] SZELAG (C. R.): A short-term forecasting algorithm for trunk demand servicing. *The Bell Technical Journal*, Vol. 61, No. 1, pp. 67 à 96, 1982.

[21] THEIL (H.): Principles of Econometrics, *John Wiley & Sons*, New York, 1971.

[22] TOME (F. M.) et CUNHA (J. A.): Traffic forecasting with a state space model. *ITC 11*, Tokyo, 1985.

[23] WONNACOTT (T. H.) et WONNACOTT (R. J.): Regression. *John Wiley & Sons*, New York, 1981.

Bibliographie

PINDYCK (R. S.) et RUBINFELD (D. F.): Econometric Models and Econometric Forecasts, *McGraw-Hill*, New York, 1981.

SASTRI (T.): A state space modelling approach for time series forecasting. *Management Science*, Vol. 31, N° 11, pp. 1451 à 1470, 1985.