



HAL
open science

La prise en compte de variables explicatives dans les modèles de séries temporelles : application à la demande de transport et au risque routier

Ruth Bergel-Hayat

► **To cite this version:**

Ruth Bergel-Hayat. La prise en compte de variables explicatives dans les modèles de séries temporelles : application à la demande de transport et au risque routier. Mathématiques générales [math.GM]. Université Paris-Est, 2008. Français. NNT : 2008PEST0203 . tel-00432051

HAL Id: tel-00432051

<https://theses.hal.science/tel-00432051>

Submitted on 13 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE PARIS-EST
ECOLE DOCTORALE ICMS

THÈSE
pour obtenir le grade de
Docteur de l'Université Paris-Est
Discipline : Mathématiques

présentée et soutenue publiquement par

Ruth Bergel-Hayat

le 25 juin 2008

**LA PRISE EN COMPTE DE VARIABLES EXPLICATIVES
DANS LES MODELES DE SERIES TEMPORELLES
APPLICATIONS A LA DEMANDE DE TRANSPORT
ET AU RISQUE ROUTIER**

Directeur de thèse
Georges Oppenheim

Jury Directeur de thèse : Georges Oppenheim
 Rapporteur : Robert Azencott
 Rapporteur : Jacqueline Pradel
 Examineur : Jean Chapelon
 Examineur : Jean Laterrasse
 Examineur : Jean-Patrick Lebacque

A la mémoire de mes parents Maurice Hayat et Herta Antonia Kaiserova

Remerciements

Mes remerciements s'adressent d'abord aux membres du jury :

A Jean Laterrasse pour l'amitié qu'il m'a faite d'examiner ce document et de présider le jury,

A Georges Oppenheim pour avoir accepté de diriger ce travail et m'avoir patiemment amenée à en formuler le questionnement,

A Robert Azencott pour m'avoir fait, de longue date, découvrir l'intérêt de l'analyse des séries chronologiques et de leurs applications, pour m'avoir aidée à dégager les axes de réflexion et avoir accepté de rapporter ce travail,

A Jacqueline Pradel pour avoir accepté de rapporter ce travail et pour m'avoir prodigué critiques justifiées et conseils avisés,

A Jean-Patrick Lebacque pour m'avoir fait profiter de son observation constructive nourrie par un intérêt déjà ancien pour l'analyse des séries chronologiques,

A Jean Chapelon pour son intérêt à se joindre au jury, et pour avoir consacré des moyens financiers afin que les connaissances en sécurité routière puissent progresser.

Ils s'adressent ensuite à celles et ceux qui ont permis que ce travail de thèse puisse aboutir :

A Bernard Girard, qui m'a continûment conseillée et soutenue dans l'effort, et auquel je dois, entre autres, ma curiosité pour les applications des analyses de séries chronologiques ;

A Yvonne Girard, qui m'a grandement facilité l'utilisation du système SAS ;

Aux étudiants et aux chercheurs non permanents, qui ont mis en pratique leurs connaissances dans le secteur des transports et de la sécurité routière, parmi lesquels Véronique Mignot, Valérie Nespoux et Cécile Vizatelle, ainsi que Mohammed Debbarh et Alexandre Depire auxquels je suis tout particulièrement redevable ;

A mes collègues du Ministère de l'Équipement :

A Christian Reynaud et Patrice Salini qui m'ont encouragée et soutenue dans les premiers exercices de projection, à Pierre Sélosse, Maurice Girault, Georges Malamoud et Jean Villette qui s'y sont associés, mes autres collègues m'excuseront de ne pas tous les citer ;

A Elvyne Février et Annick Haudebourg, ainsi qu'à Marie-Claire de Franclieu, Marie-Laure Garcin, Jean-Claude Meteyer et Olivier Noël pour avoir facilité les décisions de financement de certains de ces travaux ;

A mes collègues de l'INRETS, et en particulier :

A mes anciens collègues du Département Evaluation et Recherche en Accidentologie qui m'ont fait partager leurs connaissances en accidentologie, dont Sylvain Lassarre et Nicole Muhlrud pour nos échanges fructueux, et Jean-François Peytavin pour sa contribution à la production du document ;

A Ariane Dupont et Akli Berri du Département Economie et Sociologie des Transports pour la relecture, le soutien motivant et les conseils ;

A mes collègues du laboratoire Génie des Réseaux de Transport et Informatique Avancée : en tout premier lieu à son directeur Gérard Scemama pour m'y avoir accueillie et m'avoir laissé le temps de finaliser ce travail de thèse, et à Maurice Aron et à Sophie Midenet, en particulier, pour m'avoir motivée et aidée afin de le mener à terme ;

A mes collègues du Centre Informatique Recherche, enfin, que j'ai beaucoup sollicités ;

A mes collègues étrangers :

A Marc Gaudry pour l'intérêt qu'il a manifesté à mes recherches,

Aux doctorants et chercheurs de réseaux européens, et en particulier à mes collègues du SWOV, l'Institut national de recherche en sécurité routière des Pays-Bas, dont Frits Bijleveld et Jacques Commandeur pour nos discussions animées sur les avantages comparés des méthodes d'analyse temporelle appliquées au risque routier ;

A mes amis :

A Paul Bassat qui a bien voulu jouer le rôle de candide éclairé, et qui a rendu ce document plus accessible ;

A Cristina Howick pour sa contribution importante au travail de traduction en anglais, et dont les conseils précieux seront toujours bienvenus ;

A Sylvie Deguy et Arnauld Ménager, Elvyne Février et Nicole Muhlrاد, dont l'amitié et le soutien indéfectibles m'ont accompagnée dans les moments incertains ;

Enfin à ma famille :

A Jean-Paul qui a partagé avec moi le chemin de la thèse, pour l'attention et la patience, pour la relecture critique, et sans le soutien duquel je n'aurais pas pu mener ce travail, et à nos enfants Karen, Antoine et Samuel auxquels j'ai volé tant de temps qu'il me reste maintenant à rattraper.

Résumé

L'objet de la thèse est d'exposer une démarche méthodologique qui vise à prendre en compte, dans les modèles de séries temporelles, des effets exogènes mesurés à l'aide de variables additionnelles, et de l'illustrer par un certain nombre d'applications au secteur des transports. Dans ces applications, le pas de temps est le jour, le mois, le trimestre voire le semestre : il s'est agi de prendre en compte des effets exogènes, de nature transitoire ou de nature durable, qui se manifestent dans le court terme.

La première partie de la thèse traite de la modélisation des séries temporelles. Nous situons le cadre formel des modèles auxquels nous nous intéressons, nous exposons la démarche suivie dans le cadre des modèles ARMA avec variables explicatives, puis dans le cadre des modèles markoviens avec variables explicatives en y détaillant le cas particulier des modèles structurels.

Les deuxième et troisième parties de la thèse regroupent deux ensembles d'applications. Le premier porte sur des données de trafics, de voyageurs et de marchandises, agrégées par mode de transport ou par grande catégorie de réseau, et le second sur des données d'accidents corporels et de victimes de la circulation routière, agrégées par grande catégorie de réseau routier. La période couverte la plus large est 1970-2000.

La plupart des applications intègre la prise en compte des effets transitoires, de nature climatique et calendaire, sur la demande de transport et sur le risque routier, et nous donnons dans la thèse les premiers résultats détaillés démontrant pour la France la significativité du facteur climatique sur le bilan routier national, mesuré en nombres d'accidents corporels et de tués.

Mots-clefs

Modèles de séries temporelles. ARMA, espace état, variables explicatives, climat, calendrier, demande de transport, risque routier.

Abstract

The aim of the thesis is to set out a methodology that includes in time-series modelling exogenous effects measured by additional variables. This methodology is illustrated by a number of applications relating to transport. In these applications, time is measured in days, months, quarters and semesters (half years). We aim to take account of exogenous effects which are either transitory or durable lasting and which manifest themselves in the short term

The first part of the thesis deals with time-series modelling. We provide a typology of time-series models and place our approach within it. We describe the approach used in ARMA modelling with explanatory variables and then in state space modelling with explanatory variables, paying special attention to structural time-series modelling.

The second and third parts bring together two groups of applications. The first group considers traffic datasets, for passengers and for freight, aggregated by mode and by main network type. The second group considers numbers of road injury accidents and casualties, aggregated by main network type. The largest period covered is 1970-2000.

Most of the applications address the transitory effects on transport demand and road risk of weather and calendar factors. We provide the first detailed results that demonstrate the significance of weather factor on road safety in France, measured by numbers of injury accidents and fatalities.

Key Words

Time series modelling, ARMA, state space, explanatory variables, weather, calendar, transport demand, road risk

Table des matières

Introduction	23
Contexte	25
Problématique.....	28
Organisation de la thèse	32
Partie I : Modélisation	37
Chapitre 1. Types de modèles	39
1.1. Les différents modèles	39
1.1.1. Structure des modèles.....	39
1.1.2. Les modèles descriptifs	40
1.1.3. Les modèles explicatifs	44
1.2. Spécification économétrique retenue dans la thèse.....	46
1.2.1. Types de variables	46
1.2.2. Variables retenues	48
1.2.3. Formulation retenue	49
Chapitre 2. Modélisation ARMA.....	53
2.1. Rappels : Modélisation sans variable exogène.....	53
2.1.1. Une modélisation par étapes	53
2.1.2. Les sorties.....	60
2.2. Prise en compte des variables exogènes.....	62
2.2.1. Problématique.....	62
2.2.2. Estimation.....	62
2.3. Les sorties.....	64
2.3.1. Les paramètres du modèle.....	64
2.3.2. La prévision.....	65
Chapitre 3. Modélisation espace état	67
3.1. Rappels : Modélisation sans variable exogène.....	67
3.1.1. Représentation markovienne	67
3.1.2. Détermination du rang.....	68
3.1.3. Estimation des matrices F, K et H.....	70

3.2.	Prise en compte des variables exogènes.....	71
3.2.1.	Problématique.....	71
3.2.2.	Les trois étapes de la méthode.....	72
3.3.	Les sorties.....	73
3.3.1.	Dynamique du vecteur d'état	73
3.3.2.	Partie exogène	74
3.3.3.	La prévision.....	74
3.4.	Cas particulier : la modélisation structurelle.....	75
3.4.1.	Le modèle structurel univarié de base.....	75
3.4.2.	Le modèle structurel univarié avec variables exogènes.....	76
3.4.3.	Le modèle structurel multivarié avec variables exogènes.....	77
Partie II : Applications à la demande de transport		81
Chapitre 4. Les modèles agrégés de demande		83
4.1.	Le contexte	83
4.2.	L'approche initiale	84
4.2.1.	Une approche annuelle	84
4.2.2.	La connaissance des élasticités	84
4.3.	Des besoins nouveaux	85
4.3.1.	Une approche de court terme	85
4.3.2.	... avec variables externes.....	86
4.3.3.	...et multimodale.....	87
4.4.	Applications retenues	87
Chapitre 5. Les projections semestrielles de transport.....		89
5.1.	Objectif.....	89
5.2.	Méthode.....	89
5.3.	Résultats	91
5.3.1.	Modèles retenus.....	91
5.3.2.	Analyse des écarts	92
5.4.	Discussion	93
Chapitre 6. Les effets de nature transitoire		101
6.1.	Objectif.....	101
6.2.	Méthode.....	101
6.2.1.	Configuration calendaire	102
6.2.2.	Météorologie	103

6.3.	Applications	105
6.3.1.	Trafic de fret.....	105
6.3.2.	Circulation routière	106
6.4.	Discussion	109
Chapitre 7.	Les effets à court terme des déterminants de la demande	115
7.1.	Problématique.....	115
7.2.	Objectif.....	116
7.3.	Méthode.....	117
7.3.1.	Spécification.....	117
7.3.2.	Comparaison avec les modèles annuels	119
7.3.3.	Validation statistique et performance empirique	119
7.4.	Applications	120
7.4.1.	Trafics intérieurs de marchandises.....	120
7.4.2.	Transports interurbains de voyageurs	125
7.5.	Actualisation.....	128
7.6.	Conclusion.....	128
7.7.	Discussion et prolongements.....	129
Partie III :	Applications au risque routier.....	135
Chapitre 8.	Les modèles agrégés de risque routier	137
8.1.	Historique de la modélisation de l'insécurité routière	137
8.2.	Le cadre	141
8.2.1.	Vers des modèles explicatifs	141
8.2.2.	Le schéma de production du risque.....	141
8.3.	Applications majeures	142
8.3.1.	Modèles descriptifs : Structure déterministe ou stochastique ?	142
8.3.2.	Modèles explicatifs : quelles variables pour mesurer les facteurs de risque ?	144
8.4.	Les facteurs de nature transitoire	150
8.4.1.	Le facteur climatique.....	151
8.4.2.	Le facteur calendaire	152
8.4.3.	Applications retenues	153
Chapitre 9.	Le modèle Giboulée : un modèle de suivi à court terme.....	155
9.1.	Objectif.....	155
9.2.	Méthode.....	155

9.2.1.	Le modèle	156
9.2.2.	Options retenues	158
9.2.3.	Innovation méthodologique.....	159
9.3.	Les principaux résultats.....	161
9.3.1.	Les sorties journalières.....	161
9.3.2.	Les sorties mensuelles.....	161
9.4.	Conclusions et perspectives	162
Chapitre 10.	Le modèle RES : un modèle de court/moyen terme.....	167
10.1.	Objectif.....	167
10.2.	Problématique.....	167
10.3.	Démarche	168
10.4.	Recherche de forme fonctionnelle.....	169
10.4.1.	Objectif.....	169
10.4.2.	Méthode.....	170
10.4.3.	Résultats	172
10.5.	Le modèle étendu	178
10.5.1.	Objectif.....	178
10.5.2.	Méthode.....	178
10.5.3.	Résultats	180
10.6.	Conclusion.....	188
Chapitre 11.	Applications des modèles espace état	195
11.1.	Objectif.....	195
11.2.	Problématique.....	195
11.3.	Démarche	196
11.3.1.	Base de données	196
11.3.2.	Outils d'estimation	197
11.4.	Représentation markovienne minimale (cas bivarié).....	197
11.4.1.	Méthode.....	198
11.4.2.	Résultats	199
11.5.	Modélisation structurelle bimodale (cas bivarié).....	200
11.5.1.	Méthode.....	200
11.5.2.	Résultats	202
11.5.3.	Apport de l'approche bivariée.....	203
11.6.	Modélisation structurelle des trois niveaux du risque (cas trivarié).....	203

Liste des tableaux

Tableau 1.1: Types de modèles	52
Tableau 5.1 : Modèles ARIMA, sur 1970-1988 pour les transports interurbains de voyageurs et sur 1975-1988 pour les trafics de fret.....	93
Tableau 5.2 : Modèles ARIMA avec indicatrices, sur 1970-1988 pour les transports interurbains de voyageurs et sur 1975-1988 pour les trafics de fret.	93
Tableau 5.3: Les écarts Réalisation/Référence, absolus et relatifs, pour les trois indicateurs de déplacements de voyageurs (route, fer, air) sur 1984-1991	95
Tableau 5.4: Les écarts Réalisation/Référence, absolus et relatifs, pour les deux indicateurs de trafics de fret (route, fer), sur 1986-1991	98
Tableau 6.1 : Poids journaliers pour les trafics de fret, estimés sur 1975-1989.....	106
Tableau 6.2 : Coefficients calendaires pour la circulation routière interurbaine, estimés sur 1985-1990.....	111
Tableau 7.1 - Variables utilisées dans l'exemple des modèles de trafics de marchandises ...	120
Tableau 7.2 - Elasticités apparentes pour le modèle univarié du trafic routier de fret, sur 1980-1992.....	121
Tableau 7.3 - Elasticités apparentes pour le modèle univarié du trafic ferroviaire de fret, sur 1980-1992 (a).....	121
Tableau 7.4 - Elasticités apparentes pour le modèle univarié du trafic ferroviaire de fret, sur 1980-1992 (b).....	122
Tableau 7.5- Elasticités apparentes pour le modèle bivarié trafic routier/ trafic ferroviaire de fret, sur 1980-1992	122
Tableau 7.6 - Performances estimatives à 1 an des modèles, sur 1982-1992	124
Tableau 7.7 - Variables utilisées dans l'exemple des modèles de transports interurbains de voyageurs	125
Tableau 7.8 : Les données de trafic en 1992	131
Tableau 7.9 : La structure des modèles de demande.....	131
Tableau 7.10: Modèles de transports interurbains de voyageurs, bases trimestrielle et semestrielle, sur 1980-1999 (Sorties SAS et E-views)	132
Tableau 7.11: Modèles de trafics de marchandises, bases trimestrielle et semestrielle, sur 1980-1999 (Sorties SAS et E-views)	133

Tableau 10.1 : Modèles avec forme fonctionnelle optimale pour l'exogène principale, sur 1975-1993.....	176
Tableau 10.2 : Modèles avec forme fonctionnelle optimale pour l'exogène principale, sur 1975-1998.....	177
Tableau 10.3 : Les paramètres du modèle initial sur données filtrées, sur 1977-1999	190
Tableau 10.4 : Effets des variables climatiques, sur 1977-1999	191
Tableau 10.5 : Les paramètres du modèle complet sur données filtrées, sur 1975-1999.....	192
Tableau 10.6 : Effets des variables conjoncturelles, sur 1975-1999	193
Tableau 10.7 : Les données de sécurité routière de 1998.....	194
Tableau 10.8 : Liste des variables, unités, sources.....	194
Tableau 11.1 : Modèle markovien minimal pour les nombres mensuels d'accidents sur les routes nationales et les autoroutes concédées, sur 1975-2000 (sorties de listing MODEST).	209
Tableau 11.2 : Modèle structurel bivarié pour les nombres mensuels d'accidents sur les routes nationales et les autoroutes concédées, sur 1975-2000 (sorties de listing Ox/ SSfPack).	213
Tableau 11.3 : Modèle structurel univarié pour les nombres mensuels d'accidents sur les autoroutes concédées, sur 1975-2000 (sorties de listing Ox/ SSfPack).....	214
Tableau 11.4 : Modèle structurel univarié pour les nombres mensuels d'accidents sur les routes nationales, sur 1975-2000 (sorties de listing Ox/ SSfPack).....	214
Tableau 11.5 : Effets des variables climatiques pour 1975-2000 (sorties Ox/SSfPack).....	215
Tableau 11.6 : Modèle structurel trivarié pour les nombres mensuels de véhicule-kilomètres, d'accidents et de tués sur les autoroutes concédées, sur 1975-2000 (sorties de listing Ox/ SSfPack).....	216
Tableau 11.7 : Modèle structurel trivarié pour les nombres mensuels de véhicule-kilomètres, d'accidents pour 100 millions de véhicule-kilomètres et de tués pour 100 millions de véhicule-kilomètres, sur 1975-2000 (sorties de listing Ox/ SSfPack).	217
Tableau 11.8 : Test de significativité, sur 1975-2000 (sorties SAS).....	218

Liste des figures

Figure 5.1: La circulation routière sur le réseau national, sur 1984-1991.	96
Figure 5.2 : Le trafic de voyageurs sur le réseau principal de la SNCF, sur 1984-1991.	96
Figure 5.3 : Le trafic de passagers d’Air Inter, sur 1984-1991.	97
Figure 5.4: Le trafic routier de marchandises, sur 1986-1991.	99
Figure 5.5 : Le trafic fret de la SNCF, sur 1986-1991.	99
Figure 6.1 : Les relations trafic-calendrier, du 1 ^{er} septembre au 31 décembre 1990.	110
Figure 6.2 et Figure 6.3 : La circulation routière sur routes nationales et sur autoroutes concedées, sur 1989-1990	112
Figure 9.1 La modélisation journalière dans le modèle Giboulée.....	163
Figure 9.2. La modélisation mensuelle dans le modèle Giboulée.....	163
Figure 9.3- Nombre journalier de tués France entière : effets météorologique et calendaire, du 1er mai au 31 août 1992	164
Figure 9.4- Nombre journalier de tués France entière : données brutes et données corrigées des variations locales, du 1er mai au 31 août 1992.....	164
Figure 9.5.- Nombre journalier de tués France entière : évolution de la tendance journalière, du 1er Mai au 31 Août 1992, comparativement à celles des années précédentes.....	164
Figure 9.6.- Nombre mensuel de tués France entière : données brutes et données corrigées des variations locales, de février 1985 à décembre 1992.	165
Figure 9.7: Nombre mensuel de tués France entière : effet du mois, de février 1985 à décembre 1992.	165
Figure 9.8: Nombre mensuel de tués France entière : données désaisonnalisées et tendance, de février 1985 à décembre 1992.....	165
Figure 10.1 : Formes de la fonction de Box-Cox, pour différentes valeurs de $\lambda(\beta>0)$	175
Figure 10.2: Formes de l’élasticité, de Y par rapport à X, pour différentes valeurs de $\lambda (\beta>0)$	175
Figure 11.1 : Série brute et série ajustée des logarithmes des nombres mensuels d’accidents sur autoroutes concedées, sur 1975-2000 (sorties MODEST).	212
Figure 11.2: Série brute et série ajustée des logarithmes des nombres mensuels d’accidents sur routes nationales, sur 1975-2000 (sorties MODEST).	212

Figure 11.3: Série brute et série ajustée corrigée des exogènes des logarithmes des nombres mensuels d'accidents sur autoroutes concédées, sur 1975-2000 (sorties Ox/ SSfPack).	219
Figure 11.4: Série brute et série ajustée corrigée des exogènes des logarithmes des nombres mensuels d'accidents sur routes nationales sur 1975-2000 (sorties Ox/ SSfPack)......	219
Figure 11.5 : Séries brutes et séries ajustées corrigées des exogènes des logarithmes des nombres mensuels de véhicule-kms, d'accidents et de tués sur autoroutes concédées, sur 1975-2000 (sorties Ox/ SSfPack)......	220

Liste des annexes

Sigles utilisés.....	253
Outils informatiques utilisés.....	255
Table de Loi Normale	257
Table de distribution de T (Loi de Student)	259
Table du Khi2*	261

Introduction

Contexte

La prise en compte de facteurs explicatifs dans les modèles d'analyse de l'évolution de l'activité du secteur des transports a constitué, depuis le début des années 1980, un objectif majeur pour le Ministère chargé des Transports. Ce constat vaut de manière générale, qu'il s'agisse d'analyser l'évolution d'indicateurs de mesure des flux de trafics, ou d'indicateurs dérivés relatifs aux nuisances engendrées, comme ceux qui mesurent le risque d'accident lié au transport et sa gravité.

Des efforts ont été consacrés par les Directions de ce Ministère pour constituer des bases de données homogènes sur longue période, pour mettre en œuvre des méthodes de modélisation appropriées sur ces bases de données, et pour développer des outils logiciels dédiés à ce type de modèles. Ces efforts ont été réalisés dans le cadre de collaborations entre les organismes d'études et de recherche chargés des transports (OEST, SES et INRETS), et des structures universitaires (Universités Paris-XI et Paris-1 Panthéon-Sorbonne).

On reconnaît trois finalités aux modèles de séries temporelles : *décrire, expliquer et prévoir*. La distinction décrire/expliquer est fondamentale et sépare les modèles en deux catégories qui s'excluent. La prédiction est possible pour ces deux types de modèles.

La nécessité d'enrichir les modèles descriptifs et de rechercher des modèles explicatifs s'impose naturellement. Les applications des modèles de séries temporelles au secteur des transports qui sont présentées dans cette thèse illustrent, de manière différente pour chacune d'entre elles, l'apport de cette dimension explicative.

Les premiers travaux présentés dans la thèse ont été réalisés de 1984 à 1992 au Service d'Analyse Economique et du Plan (SAEP) du Ministère des Transports. L'optique en était *la prévision à court terme des flux de trafics, de voyageurs et de marchandises, agrégés au niveau national par mode de transport et par catégorie de réseau*, et l'objectif une meilleure connaissance de leurs fluctuations infra annuelles. L'utilisation en France des méthodes de Box et Jenkins pour l'analyse des séries temporelles, qui en était à ses débuts au moment de ces travaux, a suscité auprès des professionnels du transport un intérêt significatif, dans la mesure où l'extrapolation de tendances par application de ces méthodes à des séries d'observations mensuelles leur permettait d'anticiper à un horizon de six mois leur niveau

d'activité sans hypothèse sectorielle, et d'élaborer dès l'automne leur prévisions d'exploitation annuelle sans référence à l'environnement économique.

Pour l'administration des transports, cette approche apportait un éclairage à la prévision à 1-2 ans, réalisée pour déterminer le volume de l'activité transport pour l'année en cours et pour l'année à venir, à partir d'hypothèses d'évolution annuelle relatives au secteur et à l'environnement économique. De fait, dès 1980-1981, à la suite du second choc pétrolier, la sensibilité de la demande de transport à l'environnement s'étant modifiée, celui-ci devenant également par la suite plus incertain et difficile à prévoir, l'utilisation de modèles de prévision à base annuelle devenait insuffisante.

L'utilisation répétée, chaque trimestre, de modèles à pas de temps mensuel et l'analyse systématique des écarts entre les réalisations et les valeurs prévues le trimestre précédent montraient que la prise en compte dans les modèles des effets de nature externe qui affectent l'évolution des flux de trafic pouvait améliorer sensiblement leur prévision à court terme. Ce constat nous a conduits à améliorer ces modèles descriptifs en y introduisant des variables additionnelles, liées à des événements exceptionnels tels qu'une grève ou une période de froid prolongé, à la situation météorologique et à la configuration calendaire.

De manière plus large, la question de la cohérence de tels modèles avec des modèles déjà utilisés sur les mêmes données agrégées de trafic mais sur une base de temps annuelle en relation avec l'offre et la demande de transport, s'est naturellement posée. Cette question a donné lieu à la constitution de bases de données mensuelles, voire trimestrielles, relatives au secteur des transports et à l'environnement économique, sur longue période, sur lesquelles des modèles de type autorégressif ont été estimés. Ces travaux ont été publiés entre 1997 et 2002. Il s'agissait à la fois de dégager l'apport de ces nouveaux modèles aux modèles de régression annuelle déjà utilisés, et de déterminer la cohérence entre ces deux types de modèles à pas de temps distinct, mesurée par l'élasticité des trafics à l'offre et à la demande de transport.

Le second groupe de travaux présentés dans la thèse a été amorcé au Ministère des Transports et s'est poursuivi à l'INRETS. L'objectif était *l'analyse de l'évolution de l'insécurité routière*, agrégée au niveau national et par catégorie de réseau routier, dans une optique à *visée explicative*. Les nouveaux centres d'intérêt des gestionnaires de la sécurité routière, conformément aux orientations du Livre Blanc sur la sécurité routière paru en 1989, nous ont conduite à nous tourner vers la modélisation de l'insécurité routière avec une orientation vers

la compréhension des phénomènes générateurs d'accidents, la préparation de mesures de sécurité routière et l'estimation des impacts de ces mesures.

Dans le contexte de tendance baissière des statistiques d'accidents observée en France depuis 1972, le mouvement s'établissait à la stabilisation à la fin des années 1980 puisque le nombre annuel de tués sur la route stagnait toujours en 1990, depuis plusieurs années consécutives, au-dessus d'une barre de 10 000 morts. La question qui se posait était, pour les gestionnaires de la sécurité routière, de déterminer les raisons de cette stagnation. La réponse à cette question a été recherchée en s'appuyant sur une approche systémique de la sécurité routière, qui consiste à prendre en compte l'ensemble des facteurs déterminants du risque routier tout en évaluant les principales mesures de sécurité routière - alors que les approches antérieures se limitaient à évaluer l'efficacité des mesures réglementaires. La conception du système de sécurité routière, adoptée au niveau international, a produit la définition d'un ensemble de variables associées à des facteurs de risque (Hakim et al, 1991). Pour autant, la difficulté réside dans la mesure des facteurs explicatifs du risque routier : en l'absence d'une variable de mesure d'un facteur de risque, au niveau spatio-temporel souhaité, on ne peut qu'approximer la variable manquante.

La décennie suivante a vu la mise en place du permis à points, instauré à l'été 1992. Les effets de ce système ne pouvaient s'inscrire que dans la durée dans la mesure où la mise en place du dispositif réglementaire s'est effectuée progressivement, et se traduire par une inflexion des tendances.

De ce fait, les efforts ont porté sur la prise en compte des facteurs de risque et sur l'analyse des tendances, dans une optique de court/moyen terme. Et ce dans deux directions : le suivi conjoncturel avec une observation à court terme (avec une base de données journalières, mesurée sur une dizaine d'années, modèle Giboulée (Bergel et al, 1995), et le suivi plus explicatif (avec une base de données mensuelles, mesurée sur une vingtaine d'années, modèle TAG (Jaeger, 1998).

Les tendances de long terme sont souvent analysées sur une base annuelle - les facteurs qui jouent à long terme ne sont pas toujours mesurés en rythme infra annuel -, ce qui facilite les comparaisons au niveau national. Des modèles à tendance stochastique, souples et économes en facteurs de risque, ont été établis pour mesurer des taux de progrès de manière comparative pour une dizaine de pays européens (Lassarre, 2001).

Enfin, depuis le début de cette décennie, la volonté politique d'améliorer le bilan routier national se traduit par un renforcement significatif des contrôles des infractions, et le contrôle automatisé des vitesses est mis en place en 2003. On assiste à une réduction des vitesses moyennes sur la route et à une diminution significative du nombre de tués à compter de 2003, et au cours des années suivantes. Dans le même temps, une observation de la sécurité routière se met en place au niveau européen (SafetyNet, 2004, <http://www.erso.eu>) : il s'agit de constituer des bases de données de sécurité routière harmonisées entre les états, en relation avec des bases de données de mobilité et de comportement, sur lesquelles des méthodes sont préconisées pour une analyse agrégée de l'évolution du risque routier.

Ce second groupe de travaux présentés dans la thèse a été réalisé en réponse aux questions et aux besoins qui viennent d'être exposés. Il rassemble des applications de modèles de séries temporelles à des indicateurs de risque routier, qui sont le plus souvent les nombres d'accidents corporels et de décès, agrégés par catégorie de réseau ou sur l'ensemble du territoire national, sur une base journalière ou mensuelle. Nous avons été conduite à utiliser des modèles formellement proches de ceux utilisés pour la demande de transport mais plus riches, qui permettent d'intégrer des liens de causalité, éventuellement discrets.

Problématique

Les deux groupes de travaux qui viennent d'être rappelés¹ sont liés puisque, pour le mode routier, les flux de trafic constituent une mesure de la demande de transport, mais constituent également une mesure de l'exposition au risque - c'est-à-dire du principal facteur de risque² communément reconnu dans le cadre conceptuel retenu pour la modélisation du risque routier. Il est naturel que la modélisation de la demande de transport routier constitue la première étape - le premier niveau - d'une modélisation du risque routier. Les facteurs de demande sont aussi des facteurs de risque, et en tant que tels peuvent être explicatifs du risque d'accident et du risque d'être victime dans un accident³.

¹ D'autres travaux, qui portent sur l'analyse des flux de trafic des autres modes de transport ont été menés, sur le court ou le long terme, mais n'ont pas été présentés dans la thèse par souci de cohérence.

² Il faut entendre "du principal facteur explicatif des accidents" : par abus de langage, courant chez les modélisateurs, le nombre d'accidents est assimilé à un indicateur de risque.

³ De la même manière, il faut entendre " explicatifs des accidents et de leur gravité" : par abus de langage, le nombre d'accidents est un indicateur du risque d'accident, et le nombre de victimes est un indicateur du risque d'être victime dans un accident

Par ailleurs, le fait que la même forme de modèle ait été utilisée dans les deux groupes de travaux nous amenait à la généraliser, et à donner un cadre méthodologique pour cette modélisation.

*L'objet de la thèse est d'exposer une démarche méthodologique qui vise à prendre en compte, dans les modèles de séries temporelles, des effets exogènes mesurés à l'aide de variables additionnelles, et de l'illustrer par un certain nombre d'applications au secteur des transports. Dans ces applications, le pas de temps est le jour, le mois, le trimestre voire le semestre : il s'est agi de prendre en compte des effets exogènes, de nature transitoire ou de nature durable, qui se manifestent dans le court terme.*⁴

Les choix possibles dans la démarche portent sur la forme de modèle, sur les variables additionnelles qui mesurent les effets exogènes et sur la manière de les prendre en compte dans le modèle, et sur le pas de temps - ce dernier dépendant, dans tous les cas, de la disponibilité des données et de l'objectif fixé.

Au plan formel, la manière d'écrire les modèles de séries temporelles a évolué : l'approche basée sur les modèles de type autorégressif et s'y ramenant - modèles ARMA et ARIMA (Box et Jenkins, 1976) - a été suivie d'une approche fondée sur une modélisation structurelle, dans laquelle les composantes non observables - cycle, tendance, saisonnalité -, estimées chacune, ne sont plus déterministes mais stochastiques (Harvey, 1989). Les théories et les techniques ont évidemment progressé depuis l'approche proposée par Box et Jenkins. Il a été démontré que les filtres de différenciation devaient être maniés avec précaution : ainsi, filtrer par différences un processus à tendance déterministe conduit à une représentation non inversible, et fait courir un risque de régression fallacieuse en présence de variables additionnelles (Engle et Granger, 1987). Des tests sont donc nécessaires pour démontrer que ces évolutions théoriques ne mettent pas en cause les résultats obtenus avec des modèles intégrés.

Nous retenons dans la thèse la forme générale de la représentation markovienne (ou espace état), dans laquelle le processus observé s'exprime en fonction d'un vecteur d'état, qui peut ne

⁴ Le choix du court terme résulte ici du pas de temps retenu, et de la nature des facteurs étudiés. Une analyse des effets à long terme de certains facteurs aurait pu être menée avec les mêmes données, mais cette option n'a pas été développée dans le cadre de la thèse.

pas être observé. Cette représentation englobe les deux cas de modèles qui viennent d'être mentionnés.

On trouve dans la littérature, pour chacun de ces types de structure, des exemples d'utilisation de variables exogènes - souvent qualifiées d'explicatives et d'intervention.

Indépendamment de la spécification de base, le constat est que l'appel aux variables exogènes est plus ou moins important : seul un petit nombre de ces variables est retenu dans le cas où l'on cherche à établir une relation significative entre un facteur donné et le phénomène d'intérêt que l'on vise à modéliser, alors qu'un grand nombre de variables peuvent être utilisées si l'on vise à élaborer une approche systémique, globale.

Dans tous les cas, le nombre total de paramètres à estimer est limité en raison de considérations pratiques, qui tiennent d'une part à la disponibilité des données et aux possibilités des algorithmes utilisés, et d'autre part à la difficulté d'interpréter ces paramètres. Enfin, l'objectif de parcimonie est particulièrement approprié dans certains cas, comme celui de la modélisation vectorielle.

Nos préoccupations, qui ont donné lieu aux applications présentées dans la thèse, ont porté sur la prise en compte d'effets de nature externe qui se manifestent à court terme, sur la demande de transport et sur le risque routier. Parmi eux, on distingue les effets transitoires, qui se résorbent (qui sont généralement de nature climatique ou calendaire, mais qui peuvent aussi résulter d'un événement spécifique tel que l'instauration d'une mesure réglementaire) et les effets durables (qui se manifestent à long terme mais peuvent aussi s'observer de manière immédiate).

Nous avons essayé d'apporter des réponses aux questions suivantes :

- Dans quelle mesure les effets de la situation climatique et de la configuration calendaire, transitoires par nature, peuvent-ils être mis en évidence sur les flux de trafic ?
- Dans quelle mesure les effets des déterminants de la demande, de nature durable, peuvent-ils également être mis en évidence à court terme sur les flux de trafic ?
- Comment se manifestent, à court/moyen terme, les effets de la situation climatique et de la configuration calendaire sur les nombres d'accidents et de victimes de la circulation routière ? comment séparer leur effet direct sur le risque d'accident et sur sa gravité, et leur effet indirect par le biais de l'exposition au risque ?

- De manière plus générale, comment les effets transitoires, de nature climatique et calendaire, peuvent-ils être pris en compte dans une modélisation de la demande de transport, et dans une modélisation du système sécurité routière ?

La majorité des applications présentées intègre de fait la prise en compte d'effets transitoires de nature climatique et calendaire, que ce soit sur la demande de transport ou sur le risque routier. Un rapide aperçu des apports des modèles de séries temporelles à visée explicative qui ont été produits dans le domaine des transports, pour la demande de transport puis pour le risque routier, montre que ces effets de nature transitoire ont été peu étudiés.

La modélisation de la demande de transport par une approche économétrique⁵ se réfère principalement aujourd'hui à la théorie de la co-intégration : on prend en compte les effets d'offre et de demande, en distinguant les effets de court et de long terme. Cette approche a été menée sur données françaises sur une base de temps infra annuelle (Meyer, 1998, Lenormand, 2002), et par exemple sur des données agrégées de Grande-Bretagne sur une base de temps annuelle (Dargay, Hanly, 2002). Une importante revue des élasticités de la demande aux prix et au revenu a été produite (Goodwin, 1992) et réactualisée pour le mode routier (Goodwin et al, 2004). Dans ces travaux, les effets de nature transitoire ne sont pas pris en compte.

L'approche économétrique de la modélisation du risque routier s'inscrit dans deux courants bien différenciés.

D'une part une modélisation à visée explicative qui s'est imposée au niveau international, à partir de bases de données nationales mesurant sur une base mensuelle un très grand nombre de facteurs de risque : la démarche DRAG (pour demande routière, accident, gravité) (Gaudry et Lassarre (Eds.), 2000). Les mêmes facteurs sont retenus pour modéliser l'exposition au risque, le risque d'accident et sa gravité. D'autre part et à l'opposé, une modélisation économe en facteurs de risque, menée sur une base annuelle ou infra annuelle (Bijleveld et al, 2008), dans laquelle les tendances des indicateurs de risque sont supposées stochastiques, avec pour seul facteur de risque l'exposition au risque, mesurée par la population, ou par le nombre de véhicule-kilomètres effectués.

⁵ La référence pour une modélisation des déplacements de voyageurs sur données françaises par une approche autre qu'économétrique est le modèle MATISSE (Morellet, 2007).

Les facteurs transitoires de nature climatique et calendaire sont donc pris en compte de manière systémique dans la démarche DRAG, sur une base de temps mensuelle et sous la forme d'un petit nombre de variables qui mesurent le nombre de jours ouvrables du mois, la température et la hauteur de pluie journalières, moyennées sur le mois.

Seule une prise en compte fine du facteur climatique, déjà réalisée sur un ensemble de pays nordiques (Fridstrom et al, 1995), a été reprise dans le modèle TRULS explicatif du risque routier en Norvège (Fridstrom, dans : Gaudry, Lassarre (Eds.), 2000).

La démarche méthodologique présentée dans la thèse permet de compléter la prise en compte fine du facteur climatique qui vient d'être citée. Nous donnons les premiers résultats détaillés démontrant pour la France la significativité du facteur climatique sur la mobilité agrégée et sur le bilan routier national, mesuré par les nombres d'accidents corporels et de victimes de la circulation routière.

Organisation de la thèse

La thèse est divisée en trois parties.

La première partie traite de la modélisation des séries temporelles.

Elle est amorcée avec une présentation dans le chapitre 1 de différents types de modèles de séries temporelles, qui situe le cadre formel des modèles auxquels nous nous intéressons. Nous y donnons une définition des modèles *descriptifs* - pour lesquels la seule variable utilisée, le temps, n'est pas considérée comme explicative -, et des modèles *explicatifs* - qui peuvent être vus comme des modèles descriptifs auxquels des variables ont été ajoutées, et qui constituent donc une extension des premiers. Nous nous intéressons à deux structures de modèles bien différenciées : les modèles de décomposition d'une part, et les modèles de type ARMA (autorégressifs, ARMA et ARIMA) d'autre part. Enfin, nous donnons la spécification économétrique la plus utilisée dans la thèse, qui dans le cadre univarié est une écriture ARMA avec variables exogènes et dans le cadre vectoriel est une représentation espace état avec exogènes.

Dans le cadre de la thèse, les variables additionnelles sont supposées déterministes. Toutefois, même si certaines variables utilisées pour quantifier l'influence d'un facteur exogène sur le processus endogène à modéliser étaient considérées comme aléatoires, elles ne seraient pas

considérées comme causales : elles déterminent le processus endogène à modéliser, et ne sont pas déterminées par lui. En ce sens, nous les qualifions indifféremment *d'explicatives ou d'exogènes (exogenéité forte)*.

La démarche suivie dans le cadre des modèles ARMA avec variables explicatives est exposée dans le chapitre 2. Nous rappelons les étapes et les sorties de la modélisation ARMA considérée d'abord sans variables exogènes, puis avec variables exogènes. Dans la spécification générale retenue, c'est un processus corrigé des effets des variables exogènes, supposé stationnaire, qui est modélisé par un ARMA. Cette formulation correspond à la réalité des champs d'application étudiés, communément admise, et par ailleurs elle est parcimonieuse.

La démarche suivie dans le cadre des modèles markoviens multivariés avec variables exogènes est ensuite exposée dans le chapitre 3. Nous rappelons les étapes et sorties de la modélisation espace état d'un processus endogène vectoriel, considérée comme précédemment sans variables exogènes puis avec variables exogènes. Nous donnons des indications pour la détermination de la dimension du vecteur d'état. Comme cela a été fait pour la modélisation ARMA univariée, les variables exogènes sont introduites dans l'équation d'observation, de sorte que c'est un processus endogène corrigé des effets des variables exogènes qui est modélisé par une représentation markovienne. Nous donnons enfin le cas particulier de la modélisation structurelle que l'on obtient en fixant la dimension du vecteur d'état et les matrices du système : dans ce cas contraint, l'on est ramené à un modèle de décomposition à composantes stochastiques pour le processus endogène, qui est à tendance linéaire locale.

Les deuxième et troisième parties de la thèse regroupent les deux ensembles d'applications, à la demande multimodale de transport puis au risque routier.

Le premier groupe d'applications porte sur des données de trafics, de voyageurs et de marchandises, agrégées sur l'ensemble du territoire national par mode de transport ou par grande catégorie de réseau. La période couverte par les données de trafic, toutes applications confondues, est 1970-2000 avec une périodicité mensuelle. Concernant les données exogènes, certains agrégats de la comptabilité nationale n'ont pu être reconstitués que sur une partie de la période et avec une périodicité trimestrielle.

Nous rappelons d'abord dans le chapitre 4 le contexte et le cadre dans lesquels s'est inscrite la modélisation par séries chronologiques de la demande multimodale de transport au Ministère des Transports, depuis le début des années 1980 à la suite du second choc pétrolier. Nous décrivons l'évolution des besoins d'une analyse des trafics à très court terme en complément des analyses menées sur une base annuelle.

Nous présentons dans les chapitres suivants trois types de modélisations qui ont été successivement menées sur des indicateurs modaux de trafics, pour les déplacements de voyageurs et pour le transport de marchandises, en insistant sur les résultats obtenus pour les deux modes routier et ferroviaire qui se partagent la plus grande part des flux transportés.

Dans le chapitre 5, nous exposons les travaux menés dans le cadre d'une publication trimestrielle réalisée de 1984 à 1992, consistant à anticiper à un horizon de six mois l'évolution de l'activité du secteur des transports, mesurée à partir de quelques indicateurs phares de trafic de voyageurs et de marchandises, et ce pour différents modes de transport. L'objectif était d'extrapoler la tendance récente, avec pour objectif secondaire d'analyser les écarts constatés entre les réalisations mensuelles et les projections réalisées le trimestre précédent, considérées comme la situation de référence.

Dans le chapitre 6, nous abordons la prise en compte des effets transitoires, de nature climatique et calendaire, sur les flux de trafic. A la différence de ce qui est réalisé dans le chapitre 5 par appel à des variables indicatrices, ces effets transitoires sont traités de manière plus systématique, par appel à des variables indicatrices et à des variables continues. Nous discutons les différentes options liées à la méthode et nous donnons deux types de résultats selon que le pas de temps retenu pour la mesure des effets transitoires est le jour ou le mois.

Enfin, dans le chapitre 7, nous abordons la prise en compte des effets à court terme des déterminants de l'offre (prix, consistance des réseaux) et de la demande (production/consommation) de transport sur les flux de trafic. Il s'est agi d'intégrer, sur une base infra annuelle, les variables exogènes qui mesurent l'offre et la demande de transport dans un modèle explicatif de l'évolution des trafics, avec pour objectif secondaire de tirer des enseignements de cette approche en regard d'une approche déjà menée sur une base annuelle. L'effort a porté sur la mesure des élasticités à un an des trafics au prix du transport et à la

croissance économique, pour les deux modes ferroviaire et routier, et sur la performance de ces modèles de court terme utilisés à un horizon d'une année.

Le deuxième groupe d'applications porte sur des données d'accidents corporels et de victimes de la circulation routière, agrégées au niveau national par grande catégorie de réseau. La période couverte par ces données de risque, toutes applications confondues, est 1970-2000 avec une périodicité mensuelle. Concernant les variables exogènes, les données climatiques n'ont été constituées qu'à compter de 1975.

Nous exposons d'abord dans le chapitre 8 le cadre général de la modélisation agrégée du risque routier : historique de la modélisation, cadre formel, évolution des formulations économétriques, et problématique retenue ; puis nous présentons plusieurs types de travaux menés avec des visées différentes sur les mêmes indicateurs agrégés d'accidents corporels et de victimes.

La première application, présentée dans le chapitre 9, porte sur la construction d'un tableau de bord mensuel de l'insécurité routière, permettant une analyse au niveau du jour, réalisée par le modèle Giboulée. Ce travail répondait à une demande de la DSCR de disposer d'un suivi des nombres mensuels d'accidents corporels et de victimes de la circulation routière qui fasse la part des facteurs déterminants de l'insécurité routière et qui permette de dégager une tendance.

La seconde application, qui fait l'objet du chapitre 10, porte sur la construction d'un modèle permettant de faire de l'analyse de court/moyen terme⁶ de l'évolution des nombres d'accidents et de victimes, agrégés par catégorie de réseau. Il s'est agi de compléter la demande précédente, et d'établir un modèle à petit nombre de facteurs de risque significatifs sur une base mensuelle, qui soit à la fois explicatif et prédictif. Nous nous limitons dans la thèse à traiter sa dimension explicative. A la différence de ce qui est présenté dans le chapitre précédent, le pas de temps retenu ici est directement le mois.

Une modélisation multivariée du risque routier est développée dans le chapitre 11, avec d'une part une approche bimodale des indicateurs du risque considérés simultanément sur les routes

⁶ soit à un horizon d'un an et plus.

nationales et les autoroutes, à niveau de risque donné ; et d'autre part une approche à trois niveaux des indicateurs du risque considérés simultanément pour les trois niveaux du risque, à réseau donné. Nous retenons le cas particulier de la modélisation structurelle, et nous confirmons les estimations données dans le chapitre 10 pour les impacts sur le risque routier de deux facteurs de risque : l'exposition au risque et le facteur climatique.

La dernière application, qui est exposée dans le chapitre 12, vise à préciser la relation entre la possible anticipation des effets de l'amnistie des fautes de conduite qui accompagne l'élection présidentielle en France et l'accidentalité. L'analyse a été limitée aux statistiques de décès, et aux deux élections de 1988 et de 1995. Nous discutons les résultats obtenus par différentes méthodes statistiques utilisées sur les mêmes données.

Enfin, dans la conclusion est présentée une synthèse des travaux réalisés et des résultats obtenus. Nous soulignons d'abord, de manière transversale, les caractéristiques des travaux qui ont été exposés dans la thèse. Nous donnons ensuite les résultats obtenus, et présentons un bilan par thème applicatif, pour les modèles de demande puis pour les modèles de risque routier, et les directions de recherche.

Partie I : Modélisation

Chapitre 1.

Types de modèles

1.1. Les différents modèles

Nous reprenons en la développant la distinction initiale en trois types⁷ de modèles proposée par Gouriéroux et Montfort : les *modèles de décomposition*, les *modèles auto projectifs* et les *modèles explicatifs* (Gouriéroux, Montfort, 1990).

Les deux premiers types de modèles sont ici regroupés sous le qualificatif de *modèles descriptifs* - la seule variable exogène utilisée, le temps, n'étant pas considérée comme variable explicative -, et l'on élargit les modèles auto projectifs à un ensemble plus large, celui des modèles autorégressifs et s'y ramenant.

Les *modèles explicatifs* peuvent être vus formellement comme des modèles descriptifs auxquels des variables exogènes ont été ajoutées et constituent donc une extension des premiers.

Nous considérerons successivement ces deux types de modèles, après avoir brièvement rappelé les différentes structures de modèles existantes.

1.1.1. Structure des modèles

Nous nous limitons ici aux modèles gaussiens - le processus à modéliser étant supposé suivre une loi normale.

Les modèles de séries chronologiques peuvent être définis de différentes manières, selon le point de vue adopté.

Un processus aléatoire peut être considéré comme constitué d'un certain nombre de composantes : le cycle, la tendance, la composante saisonnière et la composante résiduelle. Alors que seul le processus à modéliser peut être observé - à partir d'une de ses réalisations -, ses composantes ne peuvent être qu'estimées à l'aide d'un modèle : ainsi, *les modèles à composantes inobservables* (ou *modèles de décomposition*) visent à produire une estimation de chacune de ces composantes, alors qu'un modèle vise habituellement à produire une estimation de la seule variable à modéliser.

⁷ On notera à ce stade que des équivalences entre certaines formes de ces modèles ont été démontrées.

Dans le cas d'une décomposition, les composantes principales (toutes les composantes, à l'exception de la composante résiduelle, qui est aléatoire par nature) peuvent être traitées comme des composantes *déterministes* ou comme des composantes aléatoires - ou encore *stochastiques*.

Enfin, le processus à modéliser peut pour différentes raisons subir une transformation avant d'être modélisé : dans le cas où cette transformation est un filtre de différenciation, on parlera de *composantes intégrées*.

En résumant ces concepts, on peut distinguer les types suivants de modèles gaussiens :

- les modèles de décomposition à composantes principales déterministes (modèle de décomposition à tendance/saisonnalité déterministe, par exemple),
- les modèles de décomposition à composantes principales stochastiques (modèle de décomposition à tendance/saisonnalité stochastique, par exemple),
- les modèles à composantes intégrées (ou encore modèles intégrés).

Cette structure de base peut être enrichie de plusieurs manières pour apporter une information additionnelle : en introduisant une référence au passé (*parties auto régressive/moyenne mobile*), et/ou à l'environnement (*variables explicatives*). Finalement, la forme elle-même du modèle, généralement linéaire par rapport aux paramètres et aux composantes, peut aussi être élargie en introduisant de la non linéarité.

La structure des modèles qui vient d'être exposée vaut pour chacune des catégories de modèles que nous abordons maintenant. Toutefois, dans chacune d'elles nous nous intéressons en particulier aux deux types de structures⁸ suivantes: les *modèles de décomposition* d'une part, et les modèles *autorégressifs et s'y ramenant* d'autre part.

1.1.2. Les modèles descriptifs

Les modèles descriptifs prennent en compte la décomposition saisonnalité/tendance/partie irrégulière de la variable à modéliser. On distingue les *modèles de décomposition* qui proposent un ajustement de chacune des trois composantes, et les modèles *de type ARMA (les modèles autorégressifs, ARMA et ARIMA)* qui proposent un ajustement de la composante irrégulière, après filtrage éventuel de la tendance et de la saisonnalité.

⁸ cf. Gouriéroux, Monfort (1990) pour une définition précise des modèles de décomposition et des modèles autoprojectifs

1.1.2.1. Les modèles de décomposition

Les différentes composantes⁹ non observables du processus se dégagent, assez naturellement : la tendance à long terme, la composante saisonnière et une composante résiduelle aléatoire.

Dans le cas d'une décomposition additive, on notera :

$$Y_t = T_t + S_t + u_t,$$

avec : T_t la tendance du processus Y_t ,

S_t la composante saisonnière, périodique,

et u_t la composante aléatoire.

La tendance est souvent pensée comme une fonction de certaines variables, qui ne sont pas toujours facilement quantifiables mais qui la déterminent - on parle aussi de ces variables comme de ses déterminants. Mais elle peut aussi être considérée comme une marche aléatoire (Harvey, 1989). La même remarque vaut pour la composante saisonnière. Ainsi, la modélisation structurelle proposée par Harvey constitue une autre forme de la décomposition précédente, dans laquelle les composantes tendancielle et saisonnière peuvent également être aléatoires. Dans ce cadre, le modèle à tendance linéaire locale (basic structural model, noté BSM et que nous donnons ici en reprenant la notation de Harvey), a été retenu pour la modélisation de données d'accidents en Grande Bretagne :

$$\begin{cases} y_t = \mu_t + \gamma_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t = \beta_{t-1} + \zeta_t \\ \gamma_t = \sum_{j=1}^{s/2} \gamma_{jt} \\ \gamma_{jt} = (\cos \frac{2\pi j}{s}) \gamma_{j,t-1} + \omega_{jt} \end{cases}$$

avec : μ_t , γ_t et ε_t les composantes tendancielle, saisonnière et résiduelle de y_t ,

β_t la pente de la composante tendancielle,

⁹ En raison de l'amplitude de la période d'étude, nous ne nous intéresserons pas ici à la composante cyclique.

et ε , η , ζ et ω_{jt} des perturbations.

Les composantes tendancielle et saisonnière peuvent être estimées par des méthodes paramétriques - on utilise souvent des fonctions polynomiales du temps -, ou par des méthodes non paramétriques - par lissage. A titre d'exemple, nous citerons en raison de sa large diffusion la méthode de décomposition du Census-X11, fondée sur l'utilisation de moyennes mobiles.

L'utilisation de moyennes mobiles pour la décomposition d'une série d'observations saisonnières en composantes saisonnière, tendancielle et irrégulière a été facilitée par la mise à disposition d'un programme de désaisonnalisation mettant en œuvre la méthode dite du Census-X11 (Dagum, 1975). Le programme initial X11-ARIMA a depuis 1975 subi quelques adaptations¹⁰, mais cette méthode de désaisonnalisation n'a pas subi de modification majeure depuis et elle reste aujourd'hui la plus utilisée pour le suivi conjoncturel de séries chronologiques à fréquence mensuelle ou trimestrielle (Ladiray, Quenneville, 2001). A titre d'exemple, dans le secteur des transports, le suivi conjoncturel est réalisé sur la base d'indicateurs mensuels représentatifs de l'activité du secteur, préalablement désaisonnalisés par cette méthode (SESP, 2008b).

1.1.2.2. Les modèles autorégressifs, ARMA et ARIMA

Dans le cas particulier où le processus Y_t est stationnaire, on utilise souvent un modèle autorégressif, ou AR, pour exprimer Y_t en fonction des valeurs qu'il a prises dans le passé, et d'une perturbation u_t . La fonction f est une forme fonctionnelle, supposée linéaire, des valeurs passées du processus, de la perturbation et de ses valeurs passées. La connaissance de la dynamique du processus permet de l'extrapoler, sous réserve qu'à l'horizon de la prévision, la structure de cette dynamique ne se soit pas modifiée. En raison de la référence au passé proche, le modèle est dit adaptatif.

Différentes écritures, qui sont équivalentes, sont utilisées pour modéliser le même processus. On retient de préférence un modèle autorégressif et moyenne mobile, ou ARMA, pour Y_t , parce que celle-ci a l'avantage de comporter un nombre restreint de paramètres. Dans ce cas,

¹⁰ Deux versions du programmes X11-ARIMA ont été livrées en 1975 et en 1988, et une première version du programme X12-ARIMA en 1998 (Ladiray, Quenneville, 2001)

Y_t est fonction des valeurs qu'il a prises dans le passé, en nombre restreint, d'une perturbation u_t et des valeurs passées de la perturbation. La fonction f donnée dans la formulation suivante ci-après est souvent supposée linéaire :

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots, u_t, u_{t-1}, \dots)$$

Dans le cas général où le processus Y_t n'est pas stationnaire, il est pratique de supposer qu'il existe un autre processus stationnaire, obtenu à partir du premier par différenciation. La classe des modèles autorégressifs et moyennes mobiles intégrés, ou ARIMA (Box et Jenkins, 1976) est de ce fait très largement utilisée pour la prévision :

$$\Phi(B)F(B)Y_t = \Theta(B)u_t$$

avec : $F(B)$ le filtre de différenciation qui a permis de stationnariser Y_t ,

$\Phi(B)$ et $\Theta(B)$ deux polynômes de l'opérateur retard B , tels que $\Phi(B)$ a des racines de module supérieur à 1 et $\Theta(B)$ a des racines de module supérieur ou égal à 1.

Les modèles ARIMA ont été popularisés suite aux publications de Box et Jenkins qui ont proposé une méthodologie pour les mettre en œuvre. Citons par exemple Khasnabis et Lyoo (1989) qui les utilisent pour modéliser et prévoir des statistiques d'accidents de camions, de 1978 à 1985, et montrent une bonne adéquation de la prévision à un horizon de deux ans.

Sur données françaises, des extrapolations à un horizon de 6 mois d'une dizaine d'indicateurs représentatifs de l'activité du secteur des transports ont été réalisées de manière périodique, par application de ces méthodes (Bergel et al, 1984 à 1992). Ces projections à court terme ont également servi de situation de référence, auxquelles les réalisations ont été comparées a posteriori, permettant ainsi une analyse des écarts importants et mettant à jour les inflexions de tendance dès qu'elles se manifestent. Après avoir été identifiés, les écarts importants et les inflexions de tendance peuvent être modélisés via l'adjonction de variables exogènes dans les modèles, comme nous le verrons dans le chapitre 5

L'utilisation de modèles ARIMA sur des indicateurs mensuels de sécurité routière n'a pas véritablement eu de finalité prédictive, mais bien explicative : l'objectif premier étant d'évaluer l'effet des grandes mesures de sécurité routière, puis dans un second temps de prendre en compte également l'ensemble des déterminants du risque routier, ce que nous détaillerons par la suite.

Nous citerons encore l'utilisation des modèles ARIMA pour affiner les modèles de décomposition : dans la méthode de désaisonnalisation du Census-X11 citée ci-dessus, l'adjonction aux données brutes d'une prévision par ARIMA à 12 mois permet de disposer d'une tendance plus robuste aux extrémités de la série.

Notons pour finir que les modèles de décomposition peuvent aussi s'écrire sous la forme de modèles ARIMA. C'est le cas pour le modèle le modèle *BSM* de Harvey, déjà cité, qui sous l'hypothèse d'indépendance des bruits ε , η , ζ et ω_{jt} est équivalent à un modèle ARIMA (Harvey, 1989).

1.1.3. Les modèles explicatifs

Dans les modèles explicatifs, quelle qu'en soit la structure de base, la variable Y_t endogène est fonction de k variables observables Z_{it} , $i=1$ à k , exogènes ou explicatives¹¹, supposées connues.

Soit $Z_t = (Z_{1t}, Z_{2t}, \dots, Z_{kt})'$ le vecteur de ces k variables exogènes et u_t la perturbation.

Le modèle est dit *statique* si les u_t sont indépendants, ou non corrélés dans le cas général où la loi du processus n'est pas gaussienne.

Le modèle est dit *dynamique* si les u_t sont auto corrélés. On suppose alors que u_t suit un modèle autorégressif, ARMA et plus généralement ARIMA.

Nous présentons maintenant successivement les modèles de décomposition avec variables explicatives, puis les modèles ARMA avec variables explicatives.

1.1.3.1. Modèles de décomposition avec variables explicatives

Sur données non saisonnières les modèles de régression lient la variable à expliquer - ou variable endogène - à des variables explicatives - ou variables exogènes. Sur données

¹¹ Exogènes ou explicatives parce qu'utilisées dans un modèle explicatif du processus endogène Y_t

saisonniers, les modèles de régression comporteront en outre des variables indicatrices de la saisonnalité (le jour, le mois, le trimestre, ...).

La méthode de désaisonnalisation de données mensuelles et trimestrielles du Census-X11 offre en fait la possibilité de prendre en compte un certain nombre de variables explicatives. Sans les citer de manière exhaustive, mentionnons le poids de certains jours du mois (Dagum, Quenneville, 1988), certains jours fériés, certains événements exceptionnels. On voit ainsi apparaître des variables explicatives liées à la saisonnalité, qui permettent à l'utilisateur de prendre en compte une composante saisonnière supplémentaire, irrégulière, complémentaire de la composante saisonnière périodique.

Enfin, le modèle structurel de Harvey avec variables explicatives (y compris d'interventions) constitue une forme d'un modèle de décomposition stochastique plus générale que le modèle structurel de base déjà mentionné, ce modèle est présenté plus avant dans le chapitre 3 et des exemples en sont donnés dans le chapitre 8.

1.1.3.2. Modèles ARMA avec variables explicatives

L'introduction de variables explicatives dans le modèle ARMA se fera sous la forme suivante

$$YC_t = Y_t - g(Z_t) = f(YC_{t-1}, YC_{t-2}, \dots, u_t, u_{t-1}, \dots)$$

où g est une fonction du vecteur de variables exogènes, qui peut prendre en compte la tendance et de saisonnalité. Ces modèles, appelés modèles ARMAX, consistent en une modélisation de la composante purement aléatoire du processus, après extraction de la composante déterministe.

Ces modèles ARMA avec variables explicatives peuvent être vus comme des modèles de régression avec perturbations ARMA, les deux formulations étant équivalentes.

Les modèles ARMA avec variables explicatives sur données mensuelles sont très fréquemment utilisés dans le domaine de la sécurité routière pour l'évaluation des effets des grandes mesures. De plus, ils intègrent généralement les effets exogènes reconnus, c'est-à-dire celui de l'exposition au risque, l'influence du climat par le biais d'une ou deux variables météorologiques, et l'impact de la configuration calendaire. On utilise des fonctions de

transfert lorsque les variables explicatives sont aléatoires et des modèles d'interventions pour évaluer les effets des mesures de sécurité routière : des exemples d'interventions seront donnés dans les applications à la demande de transport et au risque routier.

1.2. Spécification économétrique retenue dans la thèse

Nous nous intéressons ici aux modèles explicatifs dynamiques, et cherchons à modéliser simultanément la dynamique propre du processus et l'influence d'un certain nombre de facteurs exogènes, sans passer par une décomposition tendance/saisonnalité/irrégulier. La dynamique propre s'entend comme la dynamique du processus corrigé des effets exogènes.

Nous décrivons d'abord la nature des variables exogènes auxquelles nous nous intéressons, et nous donnons ensuite la spécification économétrique retenue dans le cadre de la thèse.

1.2.1. Types de variables

Les variables sont de nature déterministe (elles sont supposées connues) ou aléatoire (elles sont supposées varier suivant une loi de probabilité). Dans un modèle, la variable d'intérêt, ou variable observée (aussi qualifiée de *variable endogène*, ou *dépendante*) s'exprime en fonction de *variables exogènes*, ou *indépendantes*, et d'une perturbation. Alors que la variable endogène est toujours aléatoire, ainsi que la perturbation, les variables exogènes peuvent être déterministes ou aléatoires.

On distingue en fait trois sortes de variables exogènes : les variables d'environnement, les variables de contrôle, et les variables explicatives.

Dans le cadre de la thèse, les variables utilisées pour quantifier l'influence d'un facteur exogène sur le processus endogène à modéliser sont dans tous les cas supposées déterministes. Lorsque une variable explicative est aléatoire, la traiter comme déterministe revient à envisager la loi conditionnelle de Y_t sachant le présent et le passé de X . Si on envisage une dynamique conjointe entre Y et X , alors la notion utile sera pour X d'être fortement exogène, c'est-à-dire que la dynamique de X ne fait pas intervenir les valeurs passées de Y : on dit qu'il n'y a pas de causalité de Y vers X . On peut aussi parler d'absence de rétroaction de Y vers X .

De manière évidente, certains effets sont de nature exogène. Les exemples que nous donnons ici sont tirés du domaine des applications retenues. Une grève dans les transports, ou une situation climatique exceptionnelle telle qu'un froid prolongé, par exemple, a une incidence

immédiate et univoque sur le niveau d'activité du transport. Plus généralement, les effets conjoncturels de type climatique et calendaire, ponctuels ou répétés (cf. 1.2.2) sont de nature exogène : la situation météorologique et la configuration calendaire influencent une décision de déplacement, sans qu'une rétroaction puisse avoir lieu.

Cela peut ne pas être le cas des effets de croissance économique (production/consommation) ou de prix sur le niveau des transports des personnes et des biens. Le volume des trafics de voyageurs et celui des trafics de fret sont bien fonction de la croissance économique et des prix du transport ; en revanche, le niveau des transports a-t-il, à un horizon rapproché, une influence sur la croissance économique ou sur le prix du transport ?

Nous considérons qu'à l'exception de situations exceptionnelles (dues par exemple à une des grandes grèves que le SNCF a connues en 86/87, puis en 95 /96), une variation du niveau des transports peut effectivement ne pas avoir un tel type d'effet, et ce pour diverses raisons qui tiennent à la réalité du domaine d'application étudié. Ainsi par exemple, pour le mode ferroviaire, le niveau des tarifs pour les déplacements de voyageurs est fixé une fois par an et celui des tarifs pratiqués pour les trafics de fret, même s'ils bénéficient d'une relative liberté, sont fixés par contrat pour une durée d'au moins un an. Pour le mode routier, le péage autoroutier et le prix des carburants sont prédéterminés par les concessionnaires autoroutiers et par l'Etat.

Ajoutons enfin que les variables explicatives utilisées pour mesurer la croissance économique et le prix des transports résultent souvent d'un calcul assez complexe (interpolation, lissage) réalisé par l'INSEE ou par les opérateurs de transport, de sorte qu'il n'est pas certain qu'il soit possible de modéliser sans biais la relation d'intérêt.

Nous supposerons donc par la suite que, à l'horizon rapproché retenu (généralement le mois, ou le trimestre) dans le cadre de la thèse, les variables explicatives retenues sont causales : elles déterminent le processus endogène à modéliser et ne sont pas déterminées par lui. En ce sens, nous les qualifierons indifféremment d'*explicatives* ou d'*exogènes (exogénéité forte)*.

1.2.2. Variables retenues

Les variables exogènes peuvent être utilisées pour améliorer la performance du modèle – on modélise ainsi certaines perturbations importantes, résultants de chocs - ou pour quantifier une relation avec d'autres phénomènes liés au phénomène étudié.

On distingue les *variables indicatrices*, témoins d'un effet localisé, isolé ou répété, et les *variables de mesure* d'un phénomène dont on suppose qu'il est lié au processus Y_t et dont l'effet est permanent.

Dans la pratique, il est plus simple de modéliser d'abord les effets localisés, qui sont souvent d'un ordre de grandeur supérieur à celui des autres effets, à l'aide de variables indicatrices indiquant la présence/absence de l'effet, puis les effets permanents.

Parmi ceux-ci, l'influence du climat occupe une place à part parce qu'elle est une composante de la saisonnalité : on la modélise à l'aide de variables météorologiques descriptives d'une situation climatique.

1.2.2.1. Variables indicatrices de chocs isolés

Comme exemples de chocs isolés, citons des chocs instantanés comme une grève importante, une période de froid prolongé - qui gênera le transport ferroviaire et routier, occasionnant des reports de trafic vers un autre mode ou sur d'autres axes de transport -, ou des chocs permanents comme l'instauration d'une mesure réglementaire de sécurité routière.

Les indicatrices sont des variables d'impulsion.

1.2.2.2. Variables indicatrices de chocs répétés

Un bon exemple de chocs répétés est celui du calendrier : calendrier des jours fériés, d'une part, calendrier des congés d'autre part, générateurs de grands déplacements.

Les indicatrices sont des variables d'impulsion, de saut ou de palier (Box, Tiao, 1975, et Durbin, Koopman, 2001).

1.2.2.3. Variables météorologiques

Les variables météorologiques décrivent une situation climatique. Elles peuvent être quantitatives, mesurant des phénomènes climatiques en une centaine de stations météorologiques réparties sur le territoire national, à intervalles de temps réguliers : citons la température, la hauteur de pluie, la hauteur de neige, ou leur occurrence (absence/présence de pluie, de neige). Elles peuvent aussi être qualitatives, associées à des types de temps.

Les données fournies par Météo-France sont disponibles sur une centaine de postes d'observation répartis sur le territoire national, avec un pas de temps inférieur à la journée, qui peut être très réduit. Il est dans tous les cas nécessaire de déterminer le niveau d'agrégation spatio-temporel auquel cette information sera résumée, voire d'en sélectionner une partie.

1.2.2.4. Autres variables

Ce sont l'ensemble des autres variables, qui agissent de manière permanente. Citons le revenu des ménages, le parc de véhicules, la consistance du réseau et le prix des transports, pour modéliser un volume de trafic sur un réseau routier ; ou le volume de trafic et sa composition, la consistance du réseau, la vitesse pratiquée par les véhicules pour modéliser le risque routier sur un réseau routier.

1.2.3. Formulation retenue

Ainsi que le soulignent Azencott et al. (Azencott et al., 1997c), dont nous reprenons les notations, la dynamique d'une série temporelle se définit par la manière dont sont prises en compte la mémoire de la variable endogène, les chocs aléatoires qu'elle subit et l'influence de variables exogènes, contrôlées ou aléatoires. La diversité des écritures des modèles dynamiques univariés en économétrie conduit à les regrouper dans une formulation générale.

On retient classiquement les modèles suivants :

- les modèles de régression à retards échelonnés :
$$Y_t = \sum_{i=1}^K \Psi_i(B)Z_{i,t} + W_t$$

- les modèles ARMA :
$$\Phi(B)Y_t = \Theta(B)W_t$$

- les modèles à fonction d'intervention :
$$Y_t = \sum_{i=1}^K \Psi_i(B)Z_{it} + \Phi^{-1}(B)\Theta(B)W_t$$

- les modèles à fonction de transfert :
$$Y_t = \sum_{i=1}^K \delta_i^{-1}(B)\omega_i(B)Z_{it} + \Phi^{-1}(B)\Theta(B)W_t$$

- les modèles sous forme réduite :
$$\Phi(B)Y_t = \sum_{i=1}^K \Psi_i(B)Z_{it} + W_t$$

avec : Y la variable endogène à modéliser,

Z_i les variables exogènes, pour $i=1$ à K,

W un bruit blanc non corrélé avec le passé de Y et les Z_i ,

$\Phi(B)$, $\Psi_i(B)$, $\delta_i(B)$, $\omega_i(B)$ et $\Theta(B)$ des polynômes en B, l'opérateur retard,

tels que Φ et Θ sont unitaires, et ont des racines extérieures au cercle unité.

Dans tous ces modèles, on peut introduire des filtres de différences sur les variables, endogène ou exogènes, en travaillant sur des séries préalablement différenciées, de manière à relaxer la contrainte de stationnarité.

Les modèles précédents peuvent tous être résumés sous la forme générale des modèles ARMA avec variables exogènes (ou forme ARMAX) suivante :

$$\Phi(B)Y_t = \sum_{i=1}^K \Psi_i(B)Z_{it} + \Theta(B)W_t$$

ou sous la forme suivante :

$$\Phi(B) \left[Y_t - \sum_{i=1}^K \Phi_i(B)Z_{it} \right] = \Theta(B)W_t$$

avec, pour $i=1$ à K : $\Phi_i = \sum_{j=1}^{q_i} a_{ij}B^j$ une approximation tronquée de $\Phi^{-1}\Psi_i = \sum_{j=0}^{\infty} a_{ij}B^j$.

C'est cette formulation que nous retiendrons pour les applications présentées dans la thèse dans le cadre univarié.

Sous cette forme particulière, $\Phi(B)$ et $\Theta(B)$ décrivent la dynamique du processus corrigé des effets exogènes. Cette formulation correspond à la réalité des champs d'application étudiés, communément admise, et par ailleurs elle est parcimonieuse.

Dans cette formulation, le processus Y et les exogènes Z_i peuvent être obtenus après une transformation (filtres de différences $F(B)$ et $F_i(B)$, transformation logarithme, etc..) des données effectivement observées. En particulier, la tendance et la saisonnalité peuvent être prises en compte par les variables exogènes de telle manière que le filtrage des données ne soit plus nécessaire. Dans tous les cas, la variable endogène, éventuellement transformée et corrigée des effets exogènes : $Y_t - \sum_{i=1}^K \Phi_i(B)Z_{it}$ doit être stationnaire. On est donc amené à estimer un modèle ARMA avec variables explicatives.

Dans le cadre multivarié, on utilise la forme suivante :

$$\Phi(B)Y_t = \Psi(B)Z_t + \Theta(B)W_t$$

avec: Y et W les vecteurs endogène et de bruit blanc à m composantes,

$Z_t = (Z_{1t}, Z_{2t}, \dots, Z_{kt})'$ le vecteur des k variables exogènes, supposées connues,

et Φ , Ψ , Θ des matrices dont chaque terme est un polynôme en B

Nous avons également utilisé d'autres représentations des modèles vectoriels de type VARMA, notamment la représentation markovienne minimale et la représentation par des modèles structurels. Nous donnons ici la forme générale du modèle espace état retenue :

$$\begin{cases} Y_t = HX_t + D(B)Z_t + \varepsilon_t \\ X_{t+1} = FX_t + K\eta_t \end{cases}$$

avec Y_t le processus vectoriel à m composantes, observable,

$Z_t = (Z_{1t}, Z_{2t}, \dots, Z_{kt})'$ le vecteur des k variables exogènes, supposées connues,

X_t le vecteur d'état non observable, de dimension r ,

$F(r,r)$, $K(r,m)$ et $H(m,r)$ trois matrices à coefficients réels,

$D(B)$ une matrice (m,k) dont les termes sont des polynômes en B ,

et ε_t et η_t des erreurs de moyenne 0 et de variances respectives σ_ε^2 et σ_η^2 .

Tableau 1.1: Types de modèles

<u>Modèles descriptifs</u>	<u>Modèles explicatifs</u>
<p>Modèles de décomposition(*)</p> $Y_t = T_t + S_t + u_t$	<p>Modèles de décomposition avec variables explicatives</p> $YC_t = Y_t - g(Z_t) = T_t + S_t + u_t$
<p>Modèles autorégressifs</p> $Y_t = f(Y_{t-1}, Y_{t-2}, \dots, u_t)$	<p>Modèles autorégressifs avec variables explicatives</p> $YC_t = Y_t - g(Z_t) = f(YC_{t-1}, YC_{t-2}, \dots, u_t)$
<p>Modèles autorégressifs et moyennes mobiles</p> $Y_t = f(Y_{t-1}, Y_{t-2}, \dots, u_t, u_{t-1}, \dots)$	<p>Modèles autorégressifs et moyennes mobiles avec variables explicatives</p> $YC_t = Y_t - g(Z_t) = f(YC_{t-1}, YC_{t-2}, \dots, u_t, u_{t-1}, \dots)$
<p>et par extension : Modèles ARIMA</p>	

(*) cas d'une décomposition additive, à laquelle on peut se ramener moyennant une transformation des données.

Nota : Dans cette présentation très simplifiée, f est considérée comme linéaire, et g est une fonction d'un vecteur de variables exogènes, supposées connues..

Chapitre 2.

Modélisation ARMA

Le contenu de ce chapitre s'inspire des ouvrages de référence suivants : (Box, Jenkins, 1976), (Azencott, Dacuha-Castelle, 1984), (Brockwell, Davis, 1991 et 2002).

2.1. Rappels : Modélisation sans variable exogène

2.1.1. Une modélisation par étapes

Nous rappelons ici la modélisation par étapes utilisée pour modéliser un processus $(X_t)_{t \in \mathbb{Z}}$ à temps discret du second ordre, pour lequel on dispose d'un échantillon de données $X = (x_1, x_2, \dots, x_n)$. Ces étapes successives sont la stationnarisation préliminaire, l'identification d'un modèle ARMA(p,q) pour le processus stationnarisé, l'estimation du modèle et sa validation.

Nous rappelons ensuite les sorties que l'on peut exploiter : les paramètres du modèle qui expriment la dynamique du processus, les deux chroniques d'importance, la série estimée et la série résiduelle, et enfin la prévision.

2.1.1.1. Stationnarisation

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus à temps discret, du second ordre.

Il est stationnaire au sens large si sa moyenne, sa variance et sa structure de covariance ne dépendent pas du temps :

$$\begin{aligned} E(X_t) &= \mu \\ \text{var}(X_t) &= \sigma^2 \\ \text{cov}(X_t, X_{t+l}) &= \text{cov}(X_k, X_{k+l}) \end{aligned}$$

La première condition définit la stationnarité au premier ordre, les deux suivantes la stationnarité au second ordre.

Les processus observés couramment ne sont pas stationnaires, qu'il s'agisse de non stationnarité au premier ordre (la moyenne peut varier au cours des années et en cours d'année du fait des saisons - on parle de tendance et de saisonnalité), ou de non stationnarité au second ordre (la variance peut évoluer, par exemple avec la tendance et avec la saison).

Différentes techniques sont mises en œuvre visant à « stationnariser » de tels processus. La démarche courante consiste à stationnariser en moyenne, puis en variance.

Revenons au modèle de décomposition déjà cité, que nous utilisons maintenant pour X_t , soit :

$$X_t = T_t + S_t + u_t,$$

avec : T_t la tendance du processus X_t ,

S_t la composante saisonnière, périodique,

et u_t la composante aléatoire.

Deux manières de procéder sont possibles : estimer les deux premières composantes puis en déduire la composante résiduelle, supposée stationnaire, ou filtrer le processus par différenciation. Les travaux de Engle et Granger (1987) ont montré que ce choix dépend du fait que la composante résiduelle soit stationnaire, ou non.

Dans le premier cas, l'estimation de la tendance et de la saisonnalité peut être réalisée avec des fonctions du temps. La tendance sera une fonction polynômiale du temps - par morceaux, une exponentielle décroissante, etc., la composante saisonnière sera une combinaison des indicatrices des mois, des jours, etc.

Dans le deuxième cas, on élimine ces deux composantes en utilisant le filtre général de différenciation $F(B) = (I - B)^d (I - B^s)^D$, avec d et D deux entiers positifs ou nuls, filtre qui est couramment utilisé pour stationnariser au premier ordre le processus initial X_t , saisonnier de saison s .

Cependant, la différenciation d'un processus non stationnaire peut présenter des inconvénients majeurs. Dans le cas particulier où X_t est stationnaire autour d'une tendance (ou TS, ce qui revient au cas où u_t est stationnaire), la représentation obtenue après différenciation du processus n'est plus inversible. D'une part cela amoindrit la capacité prédictive du modèle, car la perturbation alors utilisée n'est pas l'innovation du processus initial. D'autre part la

non-inversibilité du processus rend inutilisables les méthodes de Back-Forecasting car elles utilisent explicitement cette inversibilité. Il est préférable dans un tel cas de ne pas différencier, et de neutraliser la tendance en la modélisant, et enfin de modéliser l'écart - stationnaire - à la tendance, comme dans le premier cas.

Enfin, on homogénéise la variance d'une série en appliquant aux données la transformation de Box-Cox¹² de paramètre λ (Box, Cox, 1964), définie par :

$$\begin{aligned} X_t \rightarrow X_t^{(\lambda)} &= \frac{X_t^\lambda - 1}{\lambda}, \text{ si } \lambda \neq 0 \\ X_t^{(0)} &= \text{Log } X_t \end{aligned}$$

2.1.1.2. Identification

Le processus Y_t , stationnaire, peut être modélisé par la classe des ARMA(p,q) :

$$\Phi(B)Y_t = \Theta(B)u_t,$$

avec : $\Phi(B)$ et $\Theta(B)$ deux polynômes de l'opérateur retard B , de degrés p et q ,
et $(u_t)_{t \in Z}$ un bruit blanc.

L'identification consiste à déterminer p et q .

Pour ce faire, l'on fait appel à des représentations différentes du processus.

En effet, il existe plusieurs représentations pour le même processus ARMA(p,q), qui sont la représentation canonique, la représentation moyenne mobile infinie (MA infinie) et à la représentation autorégressive infinie (AR infinie).

- La représentation ARMA canonique, unique, a les propriétés suivantes : les racines de Φ et Θ sont de module supérieur à 1, et les deux polynômes Φ et Θ sont de plus unitaires et premiers entre eux. Dans ce cas, le bruit blanc est le bruit d'innovation. Il représente la partie de Y_t indépendante de son passé à l'instant t , et on écrira alors :

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + u_t + \sum_{j=1}^q \theta_j u_{t-j},$$

avec u_t le bruit d'innovation, de variance σ^2 .

- La représentation MA infinie, qui est utilisée pour le calcul de la variance de l'erreur de prévision, s'exprime de la manière suivante à partir de l'écriture canonique :

$$Y_t = \Phi^{-1}(B)\Theta(B)u_t$$

- La représentation AR infinie, qui existe lorsque le polynôme Θ n'a pas de racine unitaire - et on dit alors que le processus est inversible - , donnée par :

$$\Phi(B)\Theta^{-1}(B)Y_t = u_t$$

Ces deux dernières représentations sont utiles quand on s'intéresse à des approximations du processus ARMA(p,q) par des processus AR purs ou MA purs finis.

L'identification d'un ARMA(p,q) est réalisée en trois étapes¹³ :

- Recherche de p_0 tel que Y_t soit approximable par un AR(p_0) ;
- Recherche de q_0 tel que Y_t soit approximable par un MA(q_0) ;
- Recherche de (p,q) tel que Y_t soit un ARMA(p,q), sachant que p est plus petit ou égal à p_0 , et q plus petit ou égal à q_0 .

¹² La transformation de Box-Cox est utilisée de préférence à une transformation puissance, en raison de ses propriétés d'invariance.

¹³ Dans le cas où le processus initial comporte une composante saisonnière, de période s, on cherchera à décomposer la dynamique dans les polynômes AR et MA, en faisant apparaître des parties qui sont relatives à la saison, qui s'expriment en fonction de l'opérateur retard B^s :

$$\Phi(B)\Phi_s(B^s)F(B)X_t = \Theta(B)\Theta_s(B^s)u_t$$

On traite alors successivement l'identification de la partie saisonnière, puis celle de la partie non saisonnière (partie de court terme), ou inversement. Lorsque cette séparation de la partie relative à la saison n'est pas possible, les valeurs de p et de q, et donc le nombre des coefficients des polynômes AR et MA d'ensemble peuvent être relativement élevés.

La recherche de p_0 et de q_0 se fait souvent par l'examen des autocorrélations $\rho(h)$ - dont on déduit les « autocorrélations partielles » $\tau(h)$, d'ordre h , associées à Y_t (algorithme de Durbin) :

$$\rho(h) = \frac{\text{cov}(Y_t, Y_{t+h})}{\text{var}(Y_t)} = \frac{\gamma(h)}{\gamma(0)}, \text{ pour } h \geq 0$$

$\tau(h)$ étant défini comme le dernier coefficient de la projection linéaire de Y_t sur ses h dernières valeurs passées.

Par exemple, dans le cas où Y_t est un AR(p_0), les autocorrélations partielles d'ordre supérieur à p_0 sont nulles, et dans le cas où Y_t est un MA(q_0), les autocorrélations d'ordre supérieur à q_0 sont nulles.

On réalise des tests d'hypothèse pour déterminer si les autocorrélations partielles d'ordre h sont nulles, à partir d'un certain ordre. Ces tests sont réalisés en utilisant une estimation de ces autocorrélations théoriques - les autocorrélations empiriques -, calculée sur un échantillon d'observations de Y_t .

2.1.1.3. Estimation

Supposons p et q connus. Il s'agit maintenant d'estimer les paramètres de la dynamique, et la variance du bruit d'innovation.

Il existe plusieurs méthodes qui permettent de disposer d'une estimation préliminaire des paramètres ϕ_i , $i=1$ à p , et θ_j , $j=1$ à q , des polynômes $\Phi(B)$ et $\Theta(B)$.

Dans le cas d'un AR pur, on utilise habituellement les équations de Yule-Walker, qui lient les paramètres ϕ_i , $i=1$ à p , aux autocorrélations $\rho(h)$, $h=1$ à p . Dans le cas d'un ARMA(p,q), l'on utilisera l'une des méthodes proposées par Box et Jenkins, ou par Rissanen-Hannan par exemple.

On réalise ensuite une estimation plus précise de ces paramètres en utilisant l'estimateur du maximum de vraisemblance, qui est convergent sous l'hypothèse gaussienne.

Dans le cas où Y_t suit une loi gaussienne, la vraisemblance d'une réalisation $Y = (y_1, y_2, \dots, y_n)$ est donnée par :

$$\text{Pr ob}(Y_1 = y_1, \dots, Y_n = y_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\det \Gamma)^{1/2}} \exp\left[-\frac{1}{2} Y' \Gamma^{-1} Y\right],$$

avec $\Gamma = E(Y Y')$

et s'exprime directement en fonction du bruit d'innovation $U = (u_1, u_2, \dots, u_n)$:

$$V(\Phi, \Theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{-n/2}} \exp\left[-\frac{1}{2\sigma^2} U'U\right]$$

Maximiser la vraisemblance par rapport aux paramètres revient à maximiser la log-vraisemblance

$$L(\Phi, \Theta, \sigma^2) = \log V = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} U'U$$

ce qui revient, lorsque n est assez grand, à minimiser la somme des carrés des résidus

$U'U = \sum_{i=1}^n u_i^2$. Dans ce cas, l'algorithme du back forecasting proposé par Box et Jenkins

fournit une approximation de la vraisemblance.

Un calcul exact de la vraisemblance peut se faire par l'algorithme du filtre de Kalman (procédure ARIMA dans SAS, method ml).

La maximisation de la vraisemblance est ensuite réalisée avec un algorithme tel que celui de Newton-Raphston.

2.1.1.4. Validation

La validation du modèle est réalisée *à l'aide de tests*.

Sont à distinguer ceux qui portent sur les propriétés du résidu de l'estimation de ceux qui portent sur la validité de chaque paramètre (test de Student).

Les tests relatifs au résidu consistent à tester l'hypothèse de bruit blanc - c'est-à-dire principalement l'hypothèse de non-corrélation (test de Ljung-Box, par exemple, ou de Ljung-Box Pierce), de normalité (test de Shapiro-Wilk, ou test non paramétrique de Kolmogorov-Smirnoff), et d'indépendance (test des « up and down »). Parmi ces propriétés, la non corrélation est fondamentale. Ainsi, si l'hypothèse de normalité n'est pas validée, le calcul de la vraisemblance peut être remis en cause, mais les estimateurs peuvent néanmoins avoir de bonnes propriétés de convergence asymptotique. En revanche, l'hypothèse de non corrélation du résidu est fondamentale, car son rejet conduit nécessairement à revoir la spécification du modèle.

En pratique, toutes les propriétés du résidu ne sont pas testées : la non corrélation et la normalité le sont, et, si elles sont validées, l'indépendance en est déduite.

Les tests de Student attestant de la significativité de chaque paramètre s'effectuent en dernier, puisqu'ils ne sont praticables que si les tests portant sur le résidu ont été concluants.

Après cette étape de validation à proprement parler, la performance empirique du modèle est évaluée *à l'aide de critères*.

Ceux-ci visent à évaluer la qualité d'ajustement du modèle, ou son pouvoir de prédiction.. Dans le premier lot citons la part de variance expliquée (R^2 ¹⁴ ou R^2 ajusté), et les différents critères d'évaluation de la pertinence des estimations, qui produisent une mesure de l'écart entre les réalisations et leurs estimations: MSE, RMSE, et le MAPRE¹⁵ très fréquemment utilisé. Dans le second lot, citons le BIC ou le AIC¹⁶, et le SBC¹⁷.

Les comparaisons entre plusieurs modèles proposés pour le même échantillon de données se feront sur la base des résultats obtenus à l'issue des tests de validation, et des critères de

¹⁴ Rappelons la part de variance expliquée ou R^2 :
$$\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

¹⁵ Le MAPRE est ici défini par : $MAPRE_R = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$ ou par $MAPRE_E = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{\hat{Y}_t} \right|$, selon que

l'on se réfère à l'écart de l'estimation à la réalisation (performance du modèle) ou à l'écart de la réalisation à l'estimation (évaluations de la réalisation)

¹⁶ Le critère d'information d'Akaike (AIC) est donné par : $-2 \ln(V_m) + 2k$, avec V_m la vraisemblance calculée au maximum

performance empirique. Une comparaison entre deux modèles emboîtés peut être réalisée en utilisant une statistique du rapport de vraisemblance, ce qui peut conduire à diminuer le nombre de paramètres d'un modèle sur-paramétré.

Dans la pratique se pose enfin, après la question de la validation du modèle, la question de sa stabilité dans le temps. La stabilité des paramètres s'étudie en comparant les estimations de différents échantillons de données couvrant des intervalles de temps différents. Pour chaque nouvel échantillon, il faut examiner comment varient les réponses aux tests de validation et les critères de performance empirique.

2.1.2. Les sorties

2.1.2.1. Les paramètres du modèle

La partie autorégressive de la représentation AR infinie rend compte de la mémoire du processus en niveau Y_t , et la partie moyenne mobile de la représentation MA infinie s'exprime à partir de l'innovation, qu'on peut interpréter comme « l'erreur de prévision à un pas » - et plus précisément comme l'écart entre réalisation et projection sur le passé (voir plus loin).

Pour la représentation ARMA(p,q) canonique, l'ensemble des p+q paramètres des polynômes $\Phi(B)$ et $\Theta(B)$ suffit à définir le passé du processus Y_t , et suffit à déterminer la prévision. Dans cette expression parcimonieuse, la prévision est déterminée à partir d'un petit nombre p de réalisations passées du processus observé et d'un petit nombre q d'erreurs de prévision passées.

2.1.2.2. La série estimée et le résidu

A ce stade, l'on dispose déjà d'une estimation, pour l'intervalle de temps sur lequel la réalisation $Y = (y_1, y_2, \dots, y_n)$ est disponible, des deux composantes orthogonales du processus Y_t , données par :

$$Y_t = P_{H_{t-1}^Y}(Y_t) + u_t$$

¹⁷ Le critère bayésien de Schwarz (SBC) est donné par : $-2 \ln(V_m) + \ln(n)k$

avec : $H_{t-1}^Y = \{Y_{t'}, t' \leq t-1\}$,

et $P_{H_{t-1}^Y}(Y_t)$ la prévision à un pas, à un instant t-1 donné, définie par projection orthogonale de Y_t sur son passé H_{t-1}^Y .

L'estimation $\hat{U} = (\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n)$ de l'innovation u_t a été fournie à l'issue de la minimisation

de la somme des carrés des résidus $U'U = \sum_{i=1}^n u_i^2$, ou par le filtre de Kalman, et l'estimation

$\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$ de la prévision à un pas a été obtenue par différence avec $Y = (y_1, y_2, \dots, y_n)$.

La variance de l'erreur de prévision est donnée par σ^2 (la variance de l'innovation). Un intervalle de confiance encadrant l'estimation, à un seuil de confiance généralement fixé à 95%, est habituellement calculé - celui-ci ne nécessite que la connaissance de σ^2 .

2.1.2.3. La prévision

De manière plus générale, la prévision à h pas, à un instant t donné, est définie par projection orthogonale de Y_{t+h} sur son passé H_t^Y . Les projections se font de proche en proche, en deux étapes, par combinaison linéaire des prévisions successives et donc en fonction des valeurs passées, sur la base des relations suivantes :

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=0}^q \theta_j u_{t-j}$$

$$P_{H_t^Y}(Y_{t+h}) = \sum_{i=1}^p P_{H_t^Y}(Y_{t+h-j}) + \sum_{j=0}^q P_{H_t^Y}(u_{t+h-j})$$

avec : $P_{H_t^Y}(Y_{t+h}) = Y_{t+h}$ pour $h \leq 0$ et $P_{H_t^Y}(u_{t+h}) = 0$ pour $h > 0$

$$P_{H_t^Y}(Y_{t+h}) = P_{H_t^Y}(P_{H_{t-1}^Y}(Y_{t+h})) \text{ pour } h > 0 \text{ et } P_{H_t^Y}(u_{t+h}) = u_{t+h} \text{ pour } h \leq 0$$

L'intervalle de confiance encadrant la prévision est calculé à partir des coefficients de l'écriture MA infinie du polynôme $\Psi = \Phi^{-1}\Theta$, et la variance de l'erreur de prévision est donnée par :

$$\sigma^2 \text{ pour la prévision à un pas, et } \sigma^2(1 + \psi_1^2 + \dots + \psi_{h-1}^2) \text{ pour la prévision à h pas.}$$

2.2. Prise en compte des variables exogènes

2.2.1. Problématique

Nous cherchons maintenant à modéliser Y_t en prenant en compte k variables exogènes Z_{it} , $i=1$ à k , supposées connues, qui vont intervenir directement sur Y_t .

Soit $Z_t = (Z_{1t}, Z_{2t}, \dots, Z_{kt})'$ le vecteur de ces k variables exogènes.

Notons YC le processus corrigé des effets exogènes, qui vérifie le système :

$$\begin{aligned} Y_t - g(Z_t) &= YC_t \\ \Phi(B)YC_t &= \Theta(B)u_t \end{aligned}$$

avec : $\Phi(B)$ et $\Theta(B)$ deux polynômes de l'opérateur retard B , de degrés p et q ,
 g une forme fonctionnelle donnée,
et u_t un bruit d'innovation de variance σ^2 , non corrélé avec Z_t .

On suppose ici que le processus Y et les exogènes Z_i peuvent avoir subi une transformation préliminaire (filtre, logarithme, etc.), de telle manière que le processus YC puisse être considéré stationnaire.

En pratique, la recherche de la forme fonctionnelle g (c'est-à-dire la recherche de la relation entre Y et chaque Z_i) est abordée directement, de manière à disposer d'une spécification de départ.

La difficulté à stationnariser Y et à identifier un modèle ARIMA pour YC tient ensuite au fait que les hypothèses habituellement faites sur le processus Y (stationnarité, loi de Y) portent maintenant sur le processus corrigé des effets exogènes YC , pour lequel on ne dispose pas encore de réalisation. On cherchera dans une étape préliminaire un estimateur convergent de YC , par exemple avec une régression avec résidus auto-corrélés.

2.2.2. Estimation

L'estimation du modèle porte à la fois sur les paramètres relatifs aux effets exogènes et sur les paramètres de la dynamique.

La formulation est celle d'une régression linéaire avec résidus ARMA :

$$\Phi(B) \left[Y_t - \sum_{i=1..K} \Phi_i(B) Z_{it} \right] = \Theta(B) u_t$$

Supposons pour faciliter l'écriture que chaque exogène Z_i agit instantanément sur Y , sans retard. Reprenons ici la notation habituellement utilisée pour la régression linéaire multiple :

$$Y_t = \beta_0 + \beta_1 Z_{1t} + \dots + \beta_k Z_{kt} + \varepsilon_t$$

$$\Phi(B) \varepsilon_t = \Theta(B) u_t$$

qui devient, sous forme matricielle :

$$Y = \beta Z + \varepsilon$$

$$\phi(B) \varepsilon = \Theta(B) u$$

avec $Y = (y_1, y_2, \dots, y_n)'$ le vecteur des n observations pour Y_t

Z la matrice des n vecteurs de variables exogènes $Z_t = (Z_{1t}, Z_{2t}, \dots, Z_{kt})'$

et $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ le vecteur des paramètres de la régression.

L'estimateur des moindres carrés ordinaires de β , donné par :

$$\hat{\beta}_{OLS} = (Z'Z)^{-1} Z'Y$$

est convergent mais n'est pas à variance minimale lorsque les ε_t sont corrélés. Il peut néanmoins déjà donner une estimation de Y si la série est assez longue.

Dans le cas où les ε_t sont corrélés, le meilleur estimateur linéaire convergent de β , l'estimateur des moindres carrés généralisés, est donné par :

$$\hat{\beta}_{GLS} = (Z' \Gamma^{-1} Z)^{-1} Z' \Gamma^{-1} Y,$$

mais son calcul nécessite de connaître $\Gamma = E(\varepsilon\varepsilon')$, c'est-à-dire Φ et Θ .

On maximise directement la vraisemblance gaussienne par rapport à l'ensemble des paramètres β , Φ , Θ et σ^2 , ce qui nécessite le calcul de l'inverse de la matrice de la dérivée seconde de la log-vraisemblance.

2.3. Les sorties

Les sorties supplémentaires résultent de la présence des paramètres des variables exogènes. Nous nous intéresserons plus particulièrement à la dynamique, à la relation entre variable endogène et variables exogènes, et à la décomposition des différents effets modélisés.

2.3.1. Les paramètres du modèle

2.3.1.1. La dynamique

Rappelons ici que la dynamique estimée est maintenant celle du processus YC corrigé des effets exogènes, supposé stationnaire.

2.3.1.2. La partie exogène

L'effet global des exogènes :

La série estimée à l'issue de la modélisation ARMA est YC_t . L'effet global des exogènes est estimé par: $\sum_{i=1..K} \hat{\Phi}_i(B)Z_{it}$, et une estimation de Y_t est obtenue par la relation :

$$\hat{Y}_t = YC_t + \sum_{i=1..K} \hat{\Phi}_i(B)Z_{it}, \text{ pour } t=1 \text{ à } n.$$

Les paramètres de la partie exogène :

Les paramètres de la partie exogène permettent d'évaluer l'influence de la correction qui a contribué à stationnariser le processus YC. L'on cherche naturellement à les interpréter pour expliciter la relation qui lie les variables exogènes à l'endogène, indépendamment de la dynamique.

Plutôt que de s'intéresser aux paramètres β_i de la variable Z_i , on s'intéresse souvent à la fonction d'élasticité correspondante, donnée par : $\frac{d(\text{Log}Y)}{d(\text{Log}Z_i)}$.

Pour des petites variations de Z_i , à un instant donné, on utilise habituellement comme formulation de l'élasticité de la variable endogène Y par rapport à une variable exogène Z_i , l'expression suivante:

$$\varepsilon_{Y/Z_i} = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta Z_i}{Z_i}} = \beta_i \frac{Z_i}{Y}.$$

Notons que dans le cas très précis où la variable endogène Y et la variable exogène Z_i ont subi une transformation logarithme, le paramètre β_i représente l'élasticité de Y par rapport à Z_i , qui est alors constante.

Remarquons pour finir que dans la pratique, on parlera d'« élasticité apparente »¹⁸ de la variable endogène Y à la variable exogène Z_i , parce que la condition d'orthogonalité entre elles des variables exogènes Z_{it} , $i=1$ à k , est rarement remplie.

2.3.2. La prévision

La prévision se fait sur la base de scénarios pour les variables exogènes - supposées connues pour la prévision.

Différents scénarios sont souvent utilisés : exogène constante, choc instantané, changement de niveau, changement de pente, et combinaisons de certains d'entre eux.

¹⁸ Par abus de langage, les modélisateurs font couramment état d'une valeur d'élasticité de la variable endogène à une des variables exogènes, sans préciser qu'il s'agit en fait d'élasticité apparente chaque fois que la condition d'orthogonalité entre les variables exogènes n'est pas remplie.

Chapitre 3.

Modélisation espace état

Dans ce chapitre, nous étendons la modélisation ARMA univariée traitée précédemment au cas multivarié d'un processus VARMA (ARMA vectoriel). Nous nous intéressons plus particulièrement à la représentation markovienne (ou espace état) minimale et à la modélisation structurelle, afin de favoriser une parcimonie quant au nombre de paramètres à estimer et de rendre l'interprétation de l'état moins difficile.

Nous nous référons aux ouvrages cités dans le chapitre 2, ainsi qu'à (Akaike, 1974), (Ruchebusch, 1978), (Desai et al, 1985), (Aoki, 1988), (Azencott et al, 1997c) pour la représentation markovienne minimale, et aux ouvrages (Harvey, 1989) et (Durbin, Koopman, 2001) pour la modélisation structurelle..

Il convient de rappeler qu'il y a eu des tentatives de décomposition des processus ARMA en tendance, saison et bruit résiduel (Hillmer, Tiao, 1982), qui conduisent à des définitions et à des résultats ambigus. Mais il ne fait aucun doute que chaque représentation markovienne, minimale ou non, comporte un vecteur d'état non observable, c'est-à-dire à composantes non directement observables. Dans le cas particulier des modèles structurels que nous traitons à la fin de ce chapitre, les composantes du vecteur d'état sont définies a priori et sont qualifiées de non observables.

3.1. Rappels : Modélisation sans variable exogène

3.1.1. Représentation markovienne

Soit $(Y_t)_{t \in Z}$ un processus vectoriel stationnaire, régulier, ARMA(p,q), d'innovation W_t ,

$$\text{avec } Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{mt})'$$

$$W_t = (W_{1t}, W_{2t}, \dots, W_{mt})'$$

$$\text{et } E(W_t W_t') = \Sigma.$$

La représentation markovienne minimale de Y consiste à introduire un vecteur X, qui peut ne pas être entièrement observé, qui prend en charge la dynamique du système suivant sous la forme d'un autorégressif vectoriel d'ordre 1 :

$$\begin{aligned} X_{t+1} &= FX_t + KW_t \\ Y_t &= HX_t + W_t \end{aligned}$$

avec $F(r,r)$, $K(r,m)$ et $H(m,r)$ trois matrices à coefficients réels,
 les valeurs propres de F de module inférieur à 1,
 r la dimension du vecteur X ,
 et Σ définie positive.

On parle aussi d'une représentation espace état : la première équation du système décrit l'évolution de l'état X , c'est l'équation d'état, et la seconde décrit l'évolution de Y en fonction de X , c'est l'équation d'observation, dans laquelle X_t est prédéterminé, puisqu'il n'est fonction que des valeurs passées des inputs $W_{s,s \leq t}$.

Il existe de multiples représentations markoviennes d'un même processus. Nous nous intéressons ici à la représentation markovienne minimale, pour laquelle la dimension de X est minimale (Ruchebusch, 1978). F est alors inversible (ou de plein rang). Cette dimension minimale ou rang de la représentation markovienne, notée r , est appelée indice de Mac Millan.

La détermination du rang de la représentation markovienne est essentielle - tout comme l'est celle des ordres p et q de la représentation ARMA -, avant l'estimation des matrices F , K et H . Nous précisons ci-après la détermination du rang, puis l'estimation des matrices F , K et H .

3.1.2. Détermination du rang

On sait que le vecteur d'état X_t appartient à l'espace vectoriel P_t des prédicteurs de Y_t à l'instant t , engendré par les coordonnées de $P_{H^Y}(Y_{n+j}), j \geq 1$, avec $H_t^Y = \{Y_{it}, t' \leq t, i \leq m\}$. La dimension de cet espace vectoriel est la taille du passé de Y_t utile pour prévoir son futur.

Comme Y_t est ARMA(p,q), la dimension r de P_t est finie, et vérifie : $r \leq m \sup(p, q)$.

Plus précisément, la dimension de X est le rang de la matrice de Hankel H , de taille infinie, définie de la manière suivante :

Soit $Y_t = \sum_{k=0}^{\infty} C_k W_{t-k}$ l'écriture moyenne mobile infinie de Y_t , avec $C_k = E(Y_t Y_{t+k}') = HF^{k-1}K$.

Notons $Y_{t+j/t} = P_{H_t^Y}(Y_{t+j})$, pour simplifier l'écriture.

On a :
$$Y_{t+j/t} = \sum_{k=j}^{\infty} C_k W_{t+j-k} = \sum_{k=0}^{\infty} C_{k+j} W_{t-k}$$

On a donc :

$$\begin{bmatrix} Y_{t+1/t} \\ \dots \\ Y_{t+j/t} \\ \dots \end{bmatrix} = \begin{bmatrix} C_1 & C_2 & \dots \\ \dots & \dots & \dots \\ C_j & C_{j+1} & \dots \\ \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} W_t \\ \dots \\ W_{t-j} \\ \dots \end{bmatrix} = H \begin{bmatrix} W_t \\ \dots \\ W_{t-j} \\ \dots \end{bmatrix}$$

Dans la pratique, la recherche du rang de H se fait de la manière suivante : on considère une matrice de Hankel finie $H_{R,R}$, avec R un rang maximum que l'on se fixe, et on cherche le rang de cette matrice finie :

$$H_{R,R} = \begin{bmatrix} C_1 & C_2 & \dots & C_R \\ C_2 & C_3 & \dots & C_{R+1} \\ \dots & \dots & \dots & \dots \\ C_R & C_{R+1} & \dots & C_{2R-1} \end{bmatrix}$$

On factorise $H_{R,R}$ sous la forme d'un produit de 3 matrices U,D,V vérifiant $H_{R,R} = UDV$, avec $U'U = V'V = I$ et $D = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_R]$. On utilise pour ce faire l'algorithme SVD (Singular Value Decomposition). Il s'agit, en examinant les valeurs propres de D, de trouver une bonne approximation de la matrice de Hankel dont le rang sera considéré comme la dimension de l'état X_t .

Deai et al (Desai et al., 1985) ont introduit et qualifié d'information mutuelle entre le passé et le futur de Y_t la quantité :

$$I(Y^+, Y^-) = -\frac{1}{2} \sum_{i=1}^n \log |1 - \sigma_i^2|$$

$$\text{avec : } Y^+ = \begin{bmatrix} Y_t \\ Y_{t+1} \\ \dots \\ \dots \end{bmatrix} \text{ et } Y^- = \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ \dots \\ \dots \end{bmatrix} \text{ pour tout } t,$$

et σ_i , $i=1$ à n , les n valeurs propres de D .

Pour chaque valeur possible du rang, on calcule cette information mutuelle ; celle-ci augmente au fur et à mesure que le nombre de valeurs propres augmente et n'augmente plus lorsque l'on dépasse la valeur du rang. Il s'agit donc de détecter un coude dans son évolution.

3.1.3. Estimation des matrices F , K et H

Soit r la valeur retenue pour le rang. On définit les matrices de variance du futur et du passé de la manière suivante :

$$R_r^+ = \begin{bmatrix} C_0 & C_{-1}' & \dots & C_{r-1}' \\ C_1 & C_0 & \dots & C_{r+1}' \\ \dots & \dots & \dots & \dots \\ C_{r-1} & \dots & \dots & C_0 \end{bmatrix} \text{ et } R_r^- = \begin{bmatrix} C_0 & C_1 & \dots & C_{r-1} \\ C_1' & C_0 & \dots & C_{r+1} \\ \dots & \dots & \dots & \dots \\ C_{r-1}' & \dots & \dots & C_0 \end{bmatrix}$$

qui admettent une décomposition triangulaire supérieure : $R_r^+ = (L_r^+)(L_r^+)'$ et $R_r^- = (L_r^-)(L_r^-)'$, où L_r^+ et L_r^- sont triangulaires inférieures et inversibles.

On prend ensuite pour estimateur du vecteur d'état : $\hat{X}_t = D_r^{-\frac{1}{2}} V' (L^-)^{-1} Y_t^-$, puis on déduit successivement : \hat{F} et $K\hat{W}$ par régression de \hat{X}_{t+1} sur \hat{X}_t , puis \hat{H} et \hat{W} par régression de \hat{Y}_t sur \hat{X}_t , et enfin \hat{K} par régression de $K\hat{W}$ sur \hat{W} .

3.2. Prise en compte des variables exogènes

3.2.1. Problématique

Soit $Z_t = (Z_{1t}, Z_{2t}, \dots, Z_{kt})'$ un vecteur de k variables exogènes, que l'on suppose connues.

La représentation markovienne de Y , avec variables exogènes, est donnée par

$$\begin{aligned} X_{t+1} &= FX_t + KW_t \\ Y_t &= HX_t + D(B)Z_t + W_t \\ E(W_t W_t') &= \Sigma \end{aligned}$$

avec : $F(r,r)$, $K(r,m)$ et $H(m,r)$ trois matrices à coefficients réels,

$D(B)$ une matrice (m,k) dont les termes sont des polynômes en B ,

r la dimension du vecteur X ,

et Σ , matrice de variance-covariance de l'innovation W_t , définie positive.

Sous cette forme, $D(B)Z_t$ représente l'effet des variables exogènes sur Y_t ¹⁹. De la même manière que cela a été fait pour la modélisation ARMA décrite dans le chapitre 2, les variables exogènes sont introduites dans l'équation d'observation.

L'estimation simultanée de F , K , H , $D(B)$ et Σ pose des difficultés dans le calcul numérique (maximisation de la vraisemblance). Pour éviter cette estimation simultanée, on fixe au départ des valeurs pour $D(B)$ et pour Σ sur la base de l'estimation VAR, et on estime le triplet F, K, H .

Une estimation initiale D_0 de la matrice $D(B)$ peut être obtenue de deux manières : par une modélisation univariée avec variables exogènes de chaque composante de Y ou par une modélisation vectorielle avec exogènes (XVAR) de Y . C'est cette seconde option qui est retenue ici.

¹⁹ Comme le soulignent Azencott et al (1997c), l'introduction des variables exogènes dans la représentation markovienne peut intervenir dans l'équation d'état, dans l'équation d'observation, ou même dans les deux équations. Or l'introduction des exogènes dans l'une ou l'autre des équations est équivalente, à un changement de base près dans l'espace des états. Mais selon qu'il s'agisse plutôt d'une fonction de transfert ou d'une intervention, l'écriture la plus parcimonieuse consisterait à placer les exogènes plutôt dans l'équation d'état ou dans l'équation d'observation. Nous retenons ici cette dernière possibilité, qui correspond bien à notre problématique : nous cherchons à modéliser l'influence des exogènes sur Y_t , et non sur la dynamique de Y_t .

Au prix d'une écriture plus contraignante, dans laquelle la spécification retenue pour chaque composante du vecteur Y_t a la même structure (forme SUR – ou Seemingly Unrelated Regression), on peut utiliser une estimation par les moindres carrés ordinaires équivalente, plus simple à mettre en œuvre : dans chaque équation figurent les mêmes variables exogènes, avec les mêmes retards. Il n'y a pas de contrainte sur les coefficients, ce qui peut induire des multi-colinéarités entre variables d'une même équation, ou d'équations différentes. Néanmoins, la convergence presque sûre de l'estimateur de la matrice de variance-covariance résiduelle Σ est assurée.

Quant à l'estimation initiale D_0 de $D(B)$, elle se déduit de la forme SUR de l'écriture vectorielle XVAR retenue pour Y , d'innovation W_t , soit :

$$A(B)Y_t = C(B)Z_t + W_t$$

et plus précisément :

$$Y_t = D(B)Z_t + A^{-1}W_t$$

avec $D(B)$ une approximation tronquée de la série entière $D^* = A^{-1}C$.

3.2.2. Les trois étapes de la méthode

La méthode comporte donc trois étapes successives.

Il s'agit d'abord de produire une estimation préliminaire des effets exogènes $D(B)$. Nous retenons une estimation résultant de la forme XVAR qui vient d'être exposée, c'est-à-dire une estimation de $\Psi^{-1}\Phi$ qu'on note D_0 . A ce stade, on dispose d'une estimation provisoire du processus $YC_t = Y_t - D(B)Z_t$, qu'on note $YC_t^0 = Y_t - D_0(B)Z_t$, et d'une estimation satisfaisante de Σ , puisque la convergence presque sûre de l'estimation de Σ est assurée.

On estime ensuite simultanément F , K et H , à $D(B)=D_0$ fixé, en s'en tenant à l'estimation de Σ dont on dispose. Il s'agit en fait, comme indiqué plus en détail en 1.2., de déterminer

d'abord le rang de la représentation markovienne minimale du processus $YC_t^0 = Y_t - D_o(B)Z_t$, puis d'estimer les matrices F, K et H de cette représentation.

Enfin, on réestime D(B) à \hat{F} , \hat{K} , \hat{H} et Σ fixés. L'objectif de cette étape est l'estimation de la matrice des effets exogènes assortie d'écart-types pour ses coefficients. On maximise pour ce faire la log-vraisemblance $L(F, K, H, D)$ en D.²⁰ La convergence presque sûre des estimateurs des paramètres a été assurée, par étapes successives.

3.3. Les sorties

3.3.1. Dynamique du vecteur d'état

Après un changement de base sur X, les composantes du vecteur transformé \tilde{X} de taille r, $\tilde{X}_t = (\tilde{X}_{1,t}, \tilde{X}_{2,t}, \dots, \tilde{X}_{r,t})'$ et les valeurs propres $\lambda_1, \lambda_2, \lambda_p$ de la matrice F, vérifient le système suivant :

$$\begin{array}{rcl} \tilde{X}_{1,t} & = & \lambda_1 \tilde{X}_{1,t-1} + \dots \\ \vdots & \vdots & \vdots \\ \tilde{X}_{p,t} & = & \lambda_p \tilde{X}_{p,t-1} + \dots \end{array}$$

Une valeur propre réelle de module proche de 1, induit un effet à long terme de la composante \tilde{X}_i , et de court terme si elle est proche de 0. A ceci s'ajoute un phénomène saisonnier de \tilde{X}_i , si λ_j est complexe, qui est à long terme si le module de λ_j est proche de 1, et de court terme s'il est proche de 0. Ceci rend difficile l'interprétation des composantes de X, qui sont des combinaisons linéaires des composantes de \tilde{X} .

L'étude des valeurs propres est révélatrice du rang. En effet, on cherche à obtenir une matrice F de rang faible mais qui prenne en compte les effets principaux. Les valeurs propres les plus faibles en module correspondent à des composantes de X dont l'effet sur la dynamique s'annule rapidement.

²⁰ On utilise l'algorithme DBFGS (Dual-Broyden-Fletcher-Glodfrab-Shanno), basé sur un calcul itératif d'une approximation de l'inverse de la matrice hessienne de la vraisemblance. On recalcule exactement les coefficients de la matrice hessienne pour obtenir les écart-types des coefficients de la matrice D(B).

3.3.2. Partie exogène

Comme indiqué plus haut, on dispose pour les coefficients des estimations et des écart-types, au maximum de vraisemblance, conditionnés par les valeurs de F, K, H et Σ , - et sans tenir compte des incertitudes sur leurs estimations.

Le calcul de la Log vraisemblance par l'algorithme de Kalman donne aussi les matrices de variances-covariances du vecteur d'état X_t , du vecteur Y_t ajusté et du vecteur Y_t prévu. Le vecteur ajusté est obtenu en utilisant le passé de Y jusqu'à l'instant t, alors que le vecteur prévu est obtenu en utilisant le passé de Y jusqu'à l'instant t-1, (cf. the filtering problem and the prediction problem, comme définis dans Brockwell, Davis, 2002).

3.3.3. La prévision

Les deux équations suivantes, du système déjà défini en 3.2.1, sont utilisées pour déterminer la prévision :

$$\begin{aligned} X_{t+1} &= FX_t + KW_t \\ Y_t &= HX_t + D(B)Z_t + W_t \end{aligned}$$

Rappelons que l'on dispose des observations pour les vecteurs Y_t et Z_t , pour $t = 0$ à t .

L'algorithme de Kalman, au moment du calcul de la vraisemblance, fournit des estimations du vecteur d'état X_t , pour $t = 1$ à $t+1$, et de l'innovation W_t , pour $t = 0$ à t .

On en déduit la prévision à h pas de Y, à l'instant t donné, par projection orthogonale de Y_{t+h} sur le passé de Y et de Z, par les relations suivantes :

$$\begin{aligned} P_{H_t^{Y,Z}}(Y_{t+1}) &= HX_{t+1} + DZ_{t+1} \\ P_{H_t^{Y,Z}}(Y_{t+2}) &= HP_{H_t^{Y,Z}}(X_{t+2}) + DZ_{t+2} = HFX_{t+1} + DZ_{t+2} \\ &\dots \\ P_{H_t^{Y,Z}}(Y_{t+h}) &= HF^{h-1}X_{t+1} + DZ_{t+h} \end{aligned}$$

avec : $H_t^{Y,Z} = \{Y_{it}, Z_{jt}, t' \leq t, i \leq m, j \leq k\}$.

3.4. Cas particulier : la modélisation structurelle

Dans le cas particulier où l'on retient une décomposition à structure stochastique constituée d'une tendance linéaire locale pour le processus observé Y , le vecteur d'état X s'exprime en fonction des composantes non observables de Y (niveau, pente, saisonnalité).

Dans ce cas, les matrices F et H sont connues, et le rang de la matrice F , ou dimension de l'état, est fixé a priori. La matrice K est la matrice Identité.

Nous donnons ci-après l'écriture matricielle du modèle à tendance linéaire locale, ou modèle structurel de base (noté BSM), donné en 1.1.2.1 pour le cas univarié; puis nous introduisons des variables explicatives, y compris d'intervention, dans la formulation. Nous généralisons l'écriture au cas multivarié.

3.4.1. Le modèle structurel univarié de base

Le modèle structurel de base est donné, en reprenant la notation de Harvey (1989)²¹, par :

$$\left\{ \begin{array}{l} y_t = \mu_t + \gamma_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + v_{t-1} + \eta_t \\ v_t = \beta_{t-1} + \zeta_t \\ \gamma_t = \sum_{j=1}^{s/2} \gamma_{jt} \\ \gamma_{jt} = \cos\left(\frac{2\pi j}{s}\right) \gamma_{j,t-1} + \omega_{jt} \end{array} \right.$$

avec : μ_t , γ_t et ε_t les composantes tendancielle, saisonnière et résiduelle de y_t ,

β_t la pente de la composante tendancielle,

et ε , η , ζ et ω_j des bruits blancs variances σ_ε^2 , σ_η^2 , σ_ζ^2 et σ_ω^2 , non corrélés entre eux.

Dans le cas particulier où la composante saisonnière est une marche aléatoire :

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t,$$

l'écriture matricielle de ce modèle est donnée par :

²¹ Nous retenons ici la notation de Harvey, largement reprise dans la littérature pour la référence aux modèles structurels

$$y_t = [1 \ 0 \ 1 \ 0 \ \dots \ 0] \alpha_t + \varepsilon_t.$$

$$\alpha_t = \begin{bmatrix} \mu_t \\ v_t \\ \gamma_t \\ \vdots \\ \gamma_{t-(s-1)} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & -1 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \alpha_{t-1} + \kappa_t,$$

avec : α_t le vecteur d'état de dimension $s + 2$ et $\kappa_t = \begin{bmatrix} \eta_t \\ \zeta_t \\ \omega_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

3.4.2. Le modèle structurel univarié avec variables exogènes

Le modèle général avec variables explicatives et d'intervention est donné par :

$$\begin{cases} y_{\mu,t} &= \mu_t + \gamma_t + \mu_t + \sum_{i=1}^K \beta_i x_{it} + \lambda \omega_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + v_{t-1} + \eta_t \\ v_t &= v_{t-1} + \zeta_t \\ \gamma_{t+1} &= - \sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t \end{cases}$$

avec : k variables explicatives $x_{it}, i = 1, \dots, k$.

les paramètres associés ... $\beta_i, i = 1, \dots, k$.

et une variable d'intervention ω_t vérifiant : $\begin{cases} \omega_t = 0, & t < \tau \\ \omega_t = 1, & t \geq \tau \end{cases}$

Pour d'autres exemples de variables d'intervention, voir (Harvey, 1989), ou (Durbin, Koopman, 2001).

On se ramène à l'écriture matricielle donnée en 3.4.1., en remplaçant dans l'équation

d'observation y_t par $y_t - \sum_{i=1}^K \beta_i x_{it} - \lambda \omega_t$, ce qui donne :

$$y_t = [1 \quad 0 \quad 1 \quad 0 \quad \dots \quad 0] x_t + [\beta_1, \dots, \beta_k, \lambda] \begin{bmatrix} x_{1t} \\ \vdots \\ x_{kt} \\ \omega_t \end{bmatrix} + \varepsilon_t$$

3.4.3. Le modèle structurel multivarié avec variables exogènes

Dans le cas général d'un processus endogène vectoriel, de dimension m , avec variables exogènes, l'écriture matricielle du modèle structurel est donnée par les deux équations d'observation et d'état, comme suit.

L'équation d'observation est donnée par :

$$y_t = \left([1 \quad 0 \quad 1 \quad 0 \quad \dots \quad 0] \otimes \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \right) \alpha_t + D(B)x_t + \varepsilon_t$$

où :

$$y_t = \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{m,t} \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{m,t} \end{bmatrix}, x_t = \begin{bmatrix} x_{1t} \\ \vdots \\ x_{kt} \end{bmatrix}$$

$$\text{et } \kappa_t = \left[\eta_{1,t}, \zeta_{1,t}, \omega_{1,t}, 0, \dots, 0, \dots, \eta_{m,t}, \zeta_{m,t}, \omega_{m,t}, 0, \dots, 0 \right]'$$

sont respectivement le vecteur de m variables endogènes. $y_{1,t}, \dots, y_{m,t}$ (ayant chacune pour niveau, pente et composante saisonnière $\mu_{j,t}, \nu_{j,t}, \gamma_{j,t}$, $j=1, \dots, m$), la perturbation de l'équation d'observation, un vecteur de k variables explicatives, et le vecteur d'état de dimension $m(s+2)$.

L'équation d'état est donnée par :

$$\alpha_t = \begin{pmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & -1 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \alpha_{t-1} + \kappa_t \end{pmatrix}$$

$$\text{où : } \alpha_t = \left[\mu_{1,t}, \nu_{1,t}, \gamma_{1,t}, \dots, \gamma_{1,t-s+1}, \dots, \mu_{m,t}, \nu_{m,t}, m_{m,t}, \gamma_{m,t}, \dots, \gamma_{m,t-s+1} \right]',$$

$$\text{et } \kappa_t = \left[\eta_{1,t}, \zeta_{1,t}, \omega_{1,t}, 0, \dots, 0, \dots, \eta_{m,t}, \zeta_{m,t}, \omega_{m,t}, 0, \dots, 0 \right]'$$

Ici \otimes désigne le produit de Kronecker et κ_t la perturbation de l'équation d'état.

Le modèle structurel peut être réécrit sous la forme scalaire suivante :

$$\begin{cases} y_{1,t} = \mu_{1,t} + \gamma_{1,t} + \sum_1^k \lambda_i x_{it} + \varepsilon_{1,t} \\ y_{2,t} = \mu_{2,t} + \gamma_{2,t} + \sum_1^k \lambda_i x_{it} + \varepsilon_{2,t} \\ \vdots \\ y_{N,t} = \mu_{N,t} + \gamma_{N,t} + \sum_1^k \lambda_i x_{it} + \varepsilon_{N,t} \end{cases}, \begin{cases} \mu_{1,t+1} = \mu_{1,t} + \beta_{1,t} + \xi_{1,t} \\ \mu_{2,t+1} = \mu_{2,t} + \beta_{2,t} + \xi_{2,t} \\ \vdots \\ \mu_{N,t+1} = \mu_{N,t} + \beta_{N,t} + \xi_{N,t} \end{cases}$$

$$\begin{cases} \beta_{1,t+1} = \beta_{1,t} + \varsigma_{1,t} \\ \beta_{2,t+1} = \beta_{2,t} + \varsigma_{2,t} \\ \vdots \\ \beta_{N,t+1} = \beta_{N,t} + \varsigma_{N,t} \end{cases}, \begin{cases} \gamma_{1,t} = - \sum_{i=1}^{s-1} \gamma_{1,t-i} + \omega_{1,t} \\ \gamma_{2,t} = - \sum_{i=1}^{s-1} \gamma_{2,t-i} + \omega_{2,t} \\ \vdots \\ \gamma_{N,t} = - \sum_{i=1}^{s-1} \gamma_{N,t-i} + \omega_{N,t} \end{cases}.$$

Les matrices de variance-covariance des deux perturbations ε_t et κ_t sont données par :

$$\text{Var}\varepsilon_t = \begin{bmatrix} \text{Var}\varepsilon_{1,t} & \text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) & \dots & \text{Cov}(\varepsilon_{1,t}, \varepsilon_{m,t}) \\ \text{Cov}(\varepsilon_{2,t}, \varepsilon_{1,t}) & \text{Var}\varepsilon_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \text{Cov}(\varepsilon_{m-1,t}, \varepsilon_{m,t}) \\ \text{Cov}(\varepsilon_{m,t}, \varepsilon_{1,t}) & \dots & \text{Cov}(\varepsilon_{m,t}, \varepsilon_{m-1,t}) & \text{Var}\varepsilon_{m,t} \end{bmatrix}$$

$$\text{Var}\kappa_t = \begin{bmatrix} B & 0 & 0 & 0 \\ 0 & C & 0 & 0 \\ 0 & 0 & D & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

avec :

$$B = \begin{bmatrix} \text{Var}\eta_{1,t} & \text{Cov}(\eta_{1,t}, \eta_{2,t}) & \dots & \text{Cov}(\eta_{1,t}, \eta_{m,t}) \\ \text{Cov}(\eta_{2,t}, \eta_{1,t}) & \text{Var}\eta_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \text{Cov}(\eta_{m-1,t}, \eta_{m,t}) \\ \text{Cov}(\eta_{m,t}, \eta_{1,t}) & \dots & \text{Cov}(\eta_{m,t}, \eta_{m-1,t}) & \text{Var}\eta_{m,t} \end{bmatrix}$$

$$C = \begin{bmatrix} \text{Var}\xi_{1,t} & \text{Cov}(\xi_{1,t}, \xi_{2,t}) & \dots & \text{Cov}(\xi_{1,t}, \xi_{m,t}) \\ \text{Cov}(\xi_{2,t}, \xi_{1,t}) & \text{Var}\xi_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \text{Cov}(\xi_{m-1,t}, \xi_{m,t}) \\ \text{Cov}(\xi_{m,t}, \xi_{1,t}) & \dots & \text{Cov}(\xi_{m,t}, \xi_{m-1,t}) & \text{Var}\xi_{m,t} \end{bmatrix}$$

et

$$D = \begin{bmatrix} \text{Var}\omega_{1,t} & \text{Cov}(\omega_{1,t}, \omega_{2,t}) & \dots & \text{Cov}(\omega_{1,t}, \omega_{m,t}) \\ \text{Cov}(\omega_{2,t}, \omega_{1,t}) & \text{Var}\omega_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \text{Cov}(\omega_{m-1,t}, \omega_{m,t}) \\ \text{Cov}(\omega_{m,t}, \omega_{1,t}) & \dots & \text{Cov}(\omega_{m,t}, \omega_{m-1,t}) & \text{Var}\omega_{m,t} \end{bmatrix}$$

Partie II : Applications à la demande de transport

Chapitre 4.

Les modèles agrégés de demande

Nous nous intéressons ici aux modèles agrégés de prévision de la demande de transport : les variables à modéliser représentatives de la demande mesurent des flux de trafic de voyageurs ou de marchandises, agrégés par mode de transport (route, fer, air, voie d'eau et voie maritime) ou par grande catégorie de réseau.

Nous présentons maintenant une rétrospective sommaire²² des modèles économétriques utilisés au Ministère de l'Équipement pour simuler l'évolution des trafics de voyageurs et de marchandises, agrégés par mode de transport ou par grande catégorie de réseau.

4.1. Le contexte

On classe les modèles de prévision selon l'horizon de la prévision, entre modèles de court terme (horizon inférieur à l'année), de moyen terme (entre 2 et 5 ans) et long terme (plus de 5 ans). En pratique, cet horizon de prévision n'est pas indépendant de la périodicité des données : rappelons que l'on ne dispose de données de trafic agrégées, par mode de transport, que sur une base mensuelle. Les modèles sur données infra-annuelles - mensuelles ou trimestrielles - sont habituellement utilisés pour la prévision à moins d'un an, et les modèles sur données annuelles pour la prévision à un an et plus d'un an. Les approches développées depuis une vingtaine d'années tendent à prendre en compte simultanément dans les modèles à fréquence infra-annuelle les mouvements de court et de long terme, dans une approche dite de co-intégration.

L'approche initiale retenue par les services d'études en transport du Ministère de l'Équipement, et mise en œuvre dans les années 1970, a consisté à prendre en compte dans les modèles les effets d'offre et de demande, avec des séries chronologiques longues à pas de temps annuel, ou supérieur à l'année.

Disposant de données mesurées sur un pas de temps plus petit, nous sommes parvenus à prendre en compte d'autres effets qui se manifestent à cette échelle - on parle alors de court

²² Nous nous limitons dans la thèse aux approches de court terme, les travaux menés à plus long terme sont présentés dans les Notes de Synthèse du SES (SESP, 2008a).

terme²³. A compter des années 80, à la suite du second choc pétrolier, des besoins nouveaux ont conduit à explorer de tels modèles avec des bases de données à périodicité plus fine. L'analyse des évolutions de court terme a d'abord été menée sans référence systématique à des effets externes - suivi conjoncturel sur le passé récent et extrapolation dans le futur proche -, puis en prenant en compte systématiquement les effets externes - de nature transitoire et durable -, puis enfin en intégrant dans la modélisation les interactions qui existent entre plusieurs indicateurs de la demande de transport.

L'approche initiale et les approches menées sur données infra annuelles pour répondre à ces besoins nouveaux sont présentées dans les sections 4.2 et 4.3 qui suivent.

4.2. L'approche initiale

4.2.1. Une approche annuelle

Les premiers modèles utilisés pour simuler l'évolution de la demande globale de transport résultaient, dans l'ensemble, d'une approche économétrique classique: modèles économétriques Log-Log²⁴, de régression linéaire multiple à pas annuel, reliant sur longue période la demande globale de transport, agrégée par mode de transport ou tous modes confondus, à ses principaux déterminants, représentatifs de l'offre et de la demande de transport (cf. les modèles de prévision annuelle des trafics de marchandises PRETRAM (Chatard et al, 1980 et 1981) et de prévision annuelle des transports de personnes PRETRAP (Taroux et Buchmüller, 1980, et Taroux, 1981) utilisés pour expliquer et pour anticiper à un horizon d'un à deux ans - année en cours et année suivante - l'évolution escomptée des flux annuels de trafic). Ces modèles se sont montrés assez robustes pour reproduire de manière satisfaisante la croissance soutenue des trafics qui a marqué les années 1970.

4.2.2. La connaissance des élasticités

La connaissance des élasticités de la demande à ses déterminants constitue un objectif à part entière, et une des sorties les plus importantes des modèles. On peut ainsi réaliser des simulations de l'évolution des trafics à l'horizon de la prévision, sans pour autant avoir à utiliser le modèle lui-même.

²³ On parle aussi de modèles de prévision à court terme pour des modèles utilisés sur des données horaires, ou journalières (il s'agit alors de prévision de trafic à proprement parler, et pas de prévision de demande).

²⁴ Sous cette forme, les paramètres de la régression sont interprétables directement, et représentent l'élasticité de l'indicateur de demande à ses déterminants.

Ainsi, des scénarios additionnels sont souvent établis en complément d'un scénario central « au fil de l'eau », sur la seule connaissance de quelques valeurs d'élasticités des trafics aux déterminants représentatifs de l'offre (prix, caractéristiques du réseau) et de la demande (production/consommation) de transport.

En transport de voyageurs, une revue des valeurs d'élasticités des transports collectifs au revenu et au prix a été produite (Goodwin, 1992), et réactualisée depuis (Goodwin et al, 2004). Dans tous les cas, les modèles établis sur données annuelles fournissent des valeurs d'élasticités à moyen ou long terme.

4.3. Des besoins nouveaux

4.3.1. Une approche de court terme

La nécessité de prendre en compte les évolutions infra annuelles est apparue à la suite du second choc pétrolier : dès 1980-1981, les modèles sur données annuelles sont devenus inaptes à anticiper l'évolution de la demande de transport à un-deux ans, la sensibilité de la demande à l'environnement s'étant modifiée, l'environnement économique devenant également par la suite plus incertain et difficile à prévoir.

Les modèles sur données mensuelles ou trimestrielles ont d'abord été utilisés sur des données corrigées des variations saisonnières, ou “ désaisonnalisées ” par application d'une méthode standard, mise en oeuvre avec le programme Census-X11 (le suivi conjoncturel de l'activité transport est réalisé au Ministère sur données mensuelles désaisonnalisées).

Cependant, les trois séries obtenues par cette méthode (données désaisonnalisées, tendance et partie résiduelle) varient avec la période retenue, notamment à ses deux extrémités, ce qui rend précaire l'analyse de la fin de période, qui est souvent la période d'intérêt. Même si l'on peut atténuer cet inconvénient avec des techniques appropriées, il est plus simple pour les prévisionnistes de modéliser des données “ brutes ” sans avoir recours à cette méthode, quitte à l'appliquer ensuite sur les prévisions à des fins d'analyse (prévision par ARIMA avant désaisonnalisation).

4.3.2. ... avec variables externes

Des modèles ARIMA sur données mensuelles ont d'abord été développés par application des méthodes de Box et Jenkins²⁵ (1976), pour simuler sur longue période et projeter à très court terme les évolutions d'une dizaine d'indicateurs du secteur, représentatifs des transports de voyageurs et de marchandises. La publication trimestrielle, à partir de 1984, de prédictions à un horizon de six mois de ces indicateurs, et l'analyse systématique des écarts observés entre les dernières réalisations et la situation de référence" (les prédictions réalisées le trimestre précédent) ont permis de cerner la nature et l'ampleur des perturbations qui affectent ces indicateurs d'activité (SAEP-OEST, 1984 à 1992).

Des travaux de recherche ont ensuite été engagés en 1990, *afin d'intégrer dans la modélisation des séries temporelles l'effet des chocs* (les perturbations importantes, identifiées, telles qu'une grève, une période de froid prolongé ou des conditions de circulation atypiques) et de manière générale *pour prendre en compte toutes les variables de nature externe qui affectent, de manière localisée dans le temps ou de manière permanente, l'évolution des indicateurs d'activité du secteur des transports* (Bergel, 1990).

Il est naturel de distinguer les types d'effets selon leur durée: d'une part les effets conjoncturels, ou encore « transitoires » qui se résorbent après une durée limitée, et d'autre part les effets durables qui se manifestent à plus long terme - même s'ils peuvent aussi s'observer de manière immédiate.

Les effets transitoires, de nature climatique ou calendaire, se manifestent sur l'ensemble des transports de voyageurs, quel qu'en soit le mode, et plus particulièrement sur la circulation routière. Les trafics de fret sont également concernés, mais dans une moindre mesure.

Les effets plus durables, qui se manifestent à plus long terme sur les trafics de voyageurs et de marchandises, résultent des caractéristiques de l'offre (prix, consistance des réseaux) et de la demande (production, consommation) de transport. L'approche retenue dans le cadre de la thèse n'est pas de ce type, et l'horizon de prévision retenu ne dépasse généralement pas un an.

²⁵ Une présentation de la méthode dite « de Box et Jenkins » et de son application à des observations mensuelles de consommations de carburant automobile en France est donnée dans (Malamoud, 1978).

4.3.3. ...et multimodale

Il existe des phénomènes de concurrence et de complémentarité entre les réseaux d'un même mode de transport, ou de différents modes de transport, qui se manifestent par des interactions entre les trafics enregistrés sur ces réseaux.

Certains de ces effets sont pris en compte par les variables exogènes introduites dans le modèle explicatif de chaque trafic ; d'autres effets ne peuvent pas être pris en compte, car non mesurables ou non mesurés, et sont intégrés dans le bruit résiduel. Les innovations de deux indicateurs de trafic, par exemple, sont alors corrélées : dans ces conditions, une modélisation simultanée des deux trafics est nécessaire. C'est l'objectif de la modélisation vectorielle.

4.4. Applications retenues

Les trois types d'applications qui sont présentées dans les chapitres suivants illustrent cette évolution des besoins, présentée en 4.2.

Les modèles conçus pour simuler l'évolution des indicateurs de trafic, sur une base journalière, mensuelle ou trimestrielle, ont permis d'intégrer successivement l'effet des chocs (chapitre 5), les effets conjoncturels ou encore « transitoires » qui se résorbent après une durée limitée (chapitre 6) et les effets qui se manifestent sur une plus longue durée mais aussi dans le court terme (chapitre 7).

Chapitre 5.

Les projections semestrielles de transport

Nous résumons ici les travaux menés dans le cadre d'une publication trimestrielle “ *Les projections semestrielles de transport* ” réalisée de 1984 à 1992 au Ministère des Transports (SAEP-OEST, 1984 à 1992), dont nous étions responsable (publication et modélisation).

La publication consistait à anticiper à un horizon de six mois l'évolution de l'activité du secteur des transports, mesurée à partir de quelques indicateurs phares. Les résultats ont par ailleurs été résumés dans des notes de synthèse (Bergel et Bergel, Sélosse dans : OEST (1987 à 1992)).

5.1. Objectif

L'objectif principal a consisté à extrapoler de manière périodique - tous les trimestres -, sur une base mensuelle et à un horizon de six mois, l'évolution d'une dizaine d'indicateurs mensuels d'activité du secteur des transports.

Un objectif secondaire a été de comparer ces projections - « la référence » - aux réalisations mensuelles, connues a posteriori : l'analyse des écarts entre les réalisations mensuelles et cette référence a permis d'identifier les principaux chocs qui affectent l'activité du secteur des transports et d'en évaluer l'intensité, et de révéler les inflexions de tendance dès qu'elles se manifestent.

5.2. Méthode

Pour chaque indicateur mensuel, un modèle ARIMA est d'abord établi sur l'échantillon de données, puis enrichi par l'adjonction de variables indicatrices pour tenir compte d'un petit nombre de valeurs extrêmes considérées comme aberrantes. Chaque indicateur est ici modélisé indépendamment des autres indicateurs.

La forme générale retenue est la suivante :

$$\Phi(B)(I - B)(I - B^{12})[X_t - \sum_{i=1}^K \beta_i I_{i,t}] = \Theta(B)\Theta_{12}(B^{12})u_t,$$

avec : X_t l'indicateur endogène,

k variables explicatives $I_{it}, i = 1, \dots, k$. vérifiant $\begin{cases} I_{it} = 1, & t = \tau_i \\ I_{it} = 0 \text{ sinon} \end{cases}$,

les paramètres associés, $\beta_i, i = 1, \dots, k$.

B et B^{12} deux opérateurs retards qui permettent de pré filtrer les données observées de leur tendance et de leur saisonnalité,

$\Phi(B)$ et $\Theta(B)$ deux polynômes de l'opérateur retard B , $\Theta_{12}(B^{12})$ un polynôme de l'opérateur retard B^{12} ,

et $(u_t)_{t \in Z}$ un bruit d'innovation, de variance σ^2 .

Chaque modèle est systématiquement réactualisé, par période de trois mois, par adjonction des dernières valeurs mensuelles connues et fournit une nouvelle extrapolation à un horizon de six mois²⁶. Les séries longues, complétées de leurs projections à six mois, sont encore désaisonnalisées par une méthode classique (méthode du Census-X11) : de fait, la série désaisonnalisée - c'est-à-dire la somme tendance et résidu - est visuellement plus parlante que la série brute. Des retournements de tendance peuvent ainsi être anticipés, quelques mois avant qu'on ne les détecte avec les techniques simples d'analyse conjoncturelle utilisées sur les données brutes (comparaison d'une réalisation mensuelle à celle du même mois de l'année précédente).

On se trouve donc disposer à chaque exercice des prévisions à 1, 2 et 3 mois réalisées trois mois auparavant²⁷. Ces prévisions ont toutefois été utilisées pour servir de référence à laquelle les réalisations mensuelles ont été comparées (ou bien : les réalisations mensuelles ont été comparées à ces prévisions).

La chronologie des chocs (les écarts importants entre la réalisation et la référence) enregistrés sur la période de publication est fournie, et leur intensité est évaluée.

Dans le cas où un écart important est constaté, on s'interroge sur les raisons de cet écart (existence d'un facteur externe non pris en compte, comme une grève..) et on introduit dans le modèle une variable indicatrice pour en tenir compte à partir de cette date.

²⁶ Cet horizon correspondant de fait à la demande des professionnels du transport, et permettait à la publication d'automne de produire une estimation de l'activité annuelle du secteur.

En résumé, le modèle est évolutif de deux manières différentes :

- par réactualisation des estimations par période trois mois, avec l'adjonction de nouvelles données,

- et par introduction de variables indicatrices en cas d'événement exceptionnel identifié.

5.3. Résultats

Nous donnons ici les résultats relatifs aux cinq principaux indicateurs de trafic²⁸, parmi la dizaine d'indicateurs de la publication.

Les trois premiers portent sur les déplacements de voyageurs, pour les modes routier, ferroviaire et aérien (la circulation interurbaine sur le réseau routier national, le trafic de voyageurs sur le réseau principal de la SNCF et le trafic de passagers d'Air Inter), et les deux derniers portent sur les trafics de fret, pour les modes routier et ferroviaire (le transport routier de marchandises et le trafic de fret de la SNCF).

5.3.1. Modèles retenus

Les modèles donnés ci-après ont été établis sur les périodes de calage 1970-1988 pour les transports de voyageurs, et 1975-1988 pour les trafics de fret (cf. tableaux 5.1 et 5.2). Les modèles sur données brutes donnés dans le tableau 5.1 n'ont pas été utilisés pour la prévision, mais permettent d'évaluer le gain obtenu avec les modèles effectivement utilisés, donnés dans le tableau 5.2. Ces derniers ont été obtenus en introduisant quelques indicatrices supplémentaires pour tenir compte d'évènements exceptionnels parfaitement identifiés²⁹. Chaque modèle est alors mieux estimé, ce qui ne change pas grand-chose à l'estimation de sa dynamique - sauf pour le modèle 2 qui en est un peu simplifié - , mais qui se traduit naturellement par une diminution de l'écart-type résiduel.

Il ressort de la comparaison de ces deux tableaux que la performance empirique des seconds, mesurée par exemple par l'écart-type résiduel, est sensiblement meilleure que celle des premiers. Ce gain varie selon l'indicateur retenu, puisqu'il est de l'ordre de 4% pour la

²⁷ qui ne sont pas comparables, puisque l'écart-type de l'erreur de prévision augmente avec l'horizon de la prévision.

²⁸ Les livraisons de carburants, les immatriculations de voitures neuves, le trafic de poids lourds sur autoroute et les trafics portuaires de marchandises diverses ont été également retenus, pour lesquels des résultats spécifiques peuvent être obtenus dans les Notes de synthèse du SES.

circulation interurbaine sur le réseau routier national, et de 20% pour le trafic de voyageurs sur le réseau principal de la SNCF.

5.3.2. Analyse des écarts

Les principaux chocs enregistrés sur l'ensemble de la période considérée - c'est-à-dire la période 1984-1991 de publication de ces projections à court terme, à laquelle s'ajoutent les années antérieures qui ont été nécessaires au calage du modèle - sont dus à une grève, à une météorologie atypique (vague de froid ou période de forte chaleur) et, de manière récurrente, à la configuration du calendrier. L'ensemble des indicateurs sont concernés, et plus particulièrement ceux qui portent sur le mode routier (cf. les tableaux 5.3 et 5.4, et les graphiques 5.1 à 5.5).

Nous donnons ici des ordres de grandeur des écarts importants.

Il faut d'abord isoler les événements de nature exceptionnelle : grève, choc pétrolier, crise du Golfe, etc., comme déjà indiqué, qui peuvent avoir une incidence particulièrement forte. Ainsi, la grève de décembre 1986-janvier 1987 de la SNCF s'est traduite par une perte de trafic de voyageurs de l'ordre de 40% et de 25% pour ces deux mois, et par une perte de trafic de fret de l'ordre de 20% chaque mois, suivie d'un report de l'ordre de 18% sur février 1987. Pour le trafic de voyageurs, aucun report significatif n'apparaît sur février, mais un report modal durant la grève sur les deux autres modes de transport et en particulier sur le mode aérien, a vraisemblablement eu lieu. Il est de fait difficile d'interpréter les écarts comme des reports modaux en l'absence d'un modèle global de la concurrence modale.

Outre ces événements exceptionnels, la circulation routière et le trafic aérien subissent les aléas climatiques : on peut citer la vague de froid de janvier 1985 (les pertes de trafic ont été de l'ordre de 10% pour la circulation routière interurbaine et de 11% pour le trafic d'Air Inter), mais aussi les périodes de chaleur atypique (on enregistre un supplément de l'ordre de 8% de la circulation routière en octobre 1986 puis à nouveau en octobre 1987, deux mois qui ont été particulièrement doux pour la saison).

C'est enfin la configuration du calendrier qui occasionne les variations de trafic les plus importantes, aux périodes des fêtes et des congés. C'est bien sûr au niveau journalier que

²⁹ et sous réserve que leur effet soit significatif - c'est-à-dire que les paramètres liés aux variables indicatrices soient significativement non nuls.

l'impact de la configuration calendaire est le plus visible, mais son impact mensuel est non négligeable : par exemple, on enregistre chaque année des écarts de l'ordre de 5 à 7% du trafic routier en mars ou en avril, selon que la fête de Pâques - et donc aussi une partie des congés scolaires - se situe en mars ou en avril (Mignot, 1991)(Bergel, Mignot, 1992).

Tableau 5.1 : Modèles ARIMA, sur 1970-1988 pour les transports interurbains de voyageurs et sur 1975-1988 pour les trafics de fret.

	période	P(B)	Q(B)	écart-type rés.
Voyageurs				
Circulation réseau national	1970-1988		$(I-0.96B+0.20B^2)(I-0.52B^{12})$ 0.07 0.07 0.06	4.62
Trafic réseau principal SNCF	1970-1988		$(I-0.67B-0.18B^2)(I-0.78B^{12})$ 0.07 0.07 0.04	218.60
Trafic d'Air Inter	1970-1988		$(I-0.55B-0.24B^2)(I-0.76B^{12})$ 0.07 0.07 0.04	21.02
Marchandises				
Trafic routier de marchandises	1975-1988		$(I-0.88B+0.15B^3)(I-0.96B^{12})$ 0.06 0.06 0.01	0.506
Trafic de fret SNCF	1975-1988	$(I+0.52B+0.41B^2)$ 0.07 0.06	$(I-0.95B^{12})$ 0.02	0.255

Tableau 5.2 : Modèles ARIMA avec indicatrices, sur 1970-1988 pour les transports interurbains de voyageurs et sur 1975-1988 pour les trafics de fret.

	période	P(B)	Q(B)	points corrigés	écart-type rés.	gain
Voyageurs						
Circulation réseau national	1970-1988		$(I-0.90B+0.15B^2)(I-0.52B^{12})$ 0.07 0.07 0.06	02/78,01/85,10/85	4.418	4,4%
Trafic réseau principal SNCF	1970-1988		$(I-0.86B)(I-0.70B^{12})$ 0.03 0.05	12/86,01/87	176.9	19,1%
Trafic d'Air Inter	1970-1988		$(I-0.55B-0.20B^2)(I-0.72B^{12})$ 0.07 0.07 0.05	11-12/79, 12/86	17.17	18,3%
Marchandises						
Trafic routier de marchandises	1975-1988		$(I-0.85B+0.14B^3)(I-0.74B^{12})$ 0.06 0.06 0.06	02/88,03/88	0.461	8,9%
Trafic de fret SNCF	1975-1988	$(I+0.65B+0.49B^2)$ 0.06 0.06	$(I-0.94B^{12})$ 0.03	12/86,01-02/87	0.223	12,6%

Nota : Dans ces tableaux, les séries d'observations ont été préalablement filtrées par $(I-B)(I-B^{12})$, de manière qu'après filtrage - c'est-à-dire élimination de la tendance et de la saisonnalité - les séries filtrées soient stationnaires.

5.4. Discussion

L'intérêt de cet exercice de projection à court terme est donc double : d'une part la production d'une extrapolation à six mois, et d'autre part l'analyse systématique, pour les trois derniers mois connus, des écarts entre les réalisations mensuelles et les projections réalisées le trimestre précédent.

On dispose ainsi d'un diagnostic qui permet, par période de trois mois, d'enrichir le modèle en tenant compte de facteurs externes nouveaux. En ce sens, le modèle est doublement adaptatif : à de nouvelles données, et à de nouveaux facteurs externes.

Les résidus importants, identifiés, sont ici systématiquement corrigés, de manière à améliorer la qualité prédictive des modèles. Les chocs, isolés ou répétés, peuvent être modélisés par l'adjonction de variables indicatrices dans les modèles sur données brutes. La volonté de compléter cette approche par l'adjonction de variables exogènes continues, supposées connues, en complément des indicatrices, s'est traduite par le développement d'un logiciel spécifique (module ARMAX du progiciel MANDRAKE³⁰). Des exemples de modélisations en sont donnés ci-après dans le chapitre 6.

³⁰ Pour la modélisation ARMAX dans le progiciel MANDRAKE, et de manière générale pour une comparaison des logiciels utilisés pour une modélisation de séries chronologiques linéaires à temps discret dans le cadre univarié, une référence utile est (Borgard, Guégan, 1996)

Tableau 5.3: Les écarts Réalisation/Référence, absolus et relatifs, pour les trois indicateurs de déplacements de voyageurs (route, fer, air) sur 1984-1991

	Réal.	Ref.	Ec.abs.	Ec.rel.	Réal.	Ref.	Ec.abs.	Ec.rel.	Réal.	Ref.	Ec.abs.	Ec.rel.
janv-84	134	133	1	0,8%	3881				435			
févr-84	124	130	-6	-4,6%	3659				422	421	1	0,2%
mars-84	141	145	-4	-2,8%	4470	4633	-163	-3,5%	454	480	-26	-5,4%
avr-84	170	172	-2	-1,2%	4285	3925	360	9,2%	472	509	-37	-7,3%
mai-84	158	161	-3	-1,9%	4274	4420	-146	-3,3%	535	527	8	1,5%
juin-84	173	172	1	0,6%	4822	4802	20	0,4%	577	570	7	1,2%
juil-84	196	201	-5	-2,5%	4830	4884	-54	-1,1%	465	505	-40	-7,9%
août-84	197	203	-6	-3,0%	4268	4291	-23	-0,5%	423	433	-10	-2,3%
sept-84	164	163	1	0,6%	4276	4020	256	6,4%	510	543	-33	-6,1%
oct-84	151	147	4	2,7%	4409	4268	141	3,3%	493	485	8	1,6%
nov-84	145	138	7	5,1%	4029	4012	17	0,4%	438	438	0	0,0%
déc-84	141	136	5	3,7%	4744	4685	59	1,3%	448	453	-5	-1,1%
janv-85	124	137	-13	-9,5%	4146	4040	106	2,6%	402	451	-49	-10,9%
févr-85	130	129	1	0,8%	3720	3859	-139	-3,6%	407	435	-28	-6,4%
mars-85	145	145	0	0,0%	4778	4753	25	0,5%	477	484	-7	-1,4%
avr-85	165	173	-8	-4,6%	4197	4067	130	3,2%	491	518	-27	-5,2%
mai-85	162	162	0	0,0%	4557	4438	119	2,7%	492	517	-25	-4,8%
juin-85	172	175	-3	-1,7%	5284	4999	285	5,7%	578	571	7	1,2%
juil-85	203	200	3	1,5%	5243	4936	307	6,2%	522	497	25	5,0%
août-85	207	201	6	3,0%	4264	4511	-247	-5,5%	452	435	17	3,9%
sept-85	172	168	4	2,4%	4180	4352	-172	-4,0%	506	541	-35	-6,5%
oct-85	166	154	12	7,8%	4227	4594	-367	-8,0%	520	518	2	0,4%
nov-85	146	146	0	0,0%	3953	4328	-375	-8,7%	474	466	8	1,7%
déc-85	143	143	0	0,0%	4818	4991	-173	-3,5%	471	475	-4	-0,8%
janv-86	143	139	4	2,9%	4162	4205	-43	-1,0%	464	474	-10	-2,1%
févr-86	129	132	-3	-2,3%	3914	3928	-14	-0,4%	422	449	-27	-6,0%
mars-86	155	147	8	5,4%	4741	4843	-102	-2,1%	507	508	-1	-0,2%
avr-86	165	171	-6	-3,5%	4344	4163	181	4,3%	536	536	0	0,0%
mai-86	175	165	10	6,1%	4592	4599	-7	-0,2%	545	538	7	1,3%
juin-86	180	177	3	1,7%	5083	5229	-146	-2,8%	618	601	17	2,8%
juil-86	218	205	13	6,3%	4938	5204	-266	-5,1%	581	536	45	8,4%
août-86	221	212	9	4,2%	4018	4410	-392	-8,9%	530	479	51	10,6%
sept-86	177	177	0	0,0%	4178	4302	-124	-2,9%	575	572	3	0,5%
oct-86	175	162	13	8,0%	4268	4567	-299	-6,5%	571	562	9	1,6%
nov-86	158	153	5	3,3%	3893	4125	-232	-5,6%	502	504	-2	-0,4%
déc-86	157	150	7	4,7%	2892	4799	-1907	-39,7%	586	515	71	13,8%
janv-87	144	148	-4	-2,7%	3101	4157	-1056	-25,4%	513	515	-2	-0,4%
févr-87	141	143	-2	-1,4%	3971	3858	113	2,9%	500	483	17	3,5%
mars-87	155	165	-10	-6,1%	4303	4714	-411	-8,7%	556	561	-5	-0,9%
avr-87	191	188	3	1,6%	4298	4185	113	2,7%	619	585	34	5,8%
mai-87	181	181	0	0,0%	4939	4520	419	9,3%	559	603	-44	-7,3%
juin-87	191	190	1	0,5%	5010	5076	-66	-1,3%	607	673	-66	-9,8%
juil-87	228	224	4	1,8%	4911	5027	-116	-2,3%	601	607	-6	-1,0%
août-87	227	227	0	0,0%	4006	4260	-254	-6,0%	563	512	51	10,0%
sept-87	189	188	1	0,5%	4270	4116	154	3,7%	655	610	45	7,4%
oct-87	183	173	10	5,8%	4340	4261	79	1,9%	630	597	33	5,5%
nov-87	167,5	166,2	1,3	0,8%	3995	3974	21	0,5%	553,4	553,4	0	0,0%
déc-87	163,9	164,9	-1	-0,6%	4781	4781	0	0,0%	627,2	567,5	59,7	10,5%
janv-88	161,8	161,4	0,4	0,2%	4060	4106	-46	-1,1%	545	567,2	-22,2	-3,9%
févr-88	153,7	150,6	3,1	2,1%	4100	3857	243	6,3%	547,3	529,7	17,6	3,3%
mars-88	163,8	168,1	-4,3	-2,6%	4720	4670	50	1,1%	595,2	622,9	-27,7	-4,4%
avr-88	195,6	198,5	-2,9	-1,5%	4480	4326	154	3,6%	611,7	660,7	-49	-7,4%
mai-88	191,8	191,3	0,5	0,3%	4780	4513	267	5,9%	667,4	668,7	-1,3	-0,2%
juin-88	194,6	199,5	-4,9	-2,5%	5301	5161	140	2,7%	670,7	731,1	-60,4	-8,3%
juil-88	243,2	236	7,2	3,1%	5020	5096	-76	-1,5%	646	655,2	-9,2	-1,4%
août-88	233,6	236,7	-3,1	-1,3%	4200	4373	-173	-4,0%	589,8	587,3	2,5	0,4%
sept-88	198,6	197,5	1,1	0,6%	4400	4315	85	2,0%	688,9	674,9	14	2,1%
oct-88	195,9	189,7	6,2	3,3%	4380	4426	-46	-1,0%	695	663	32	4,8%
nov-88	175,9	176,7	-0,8	-0,5%	4185	4135	50	1,2%	624,2	591,4	32,8	5,5%
déc-88	171,6	174,4	-2,8	-1,6%	4705	4929	-224	-4,5%	649,6	624,7	24,9	4,0%
janv-89	172,9	171,6	1,3	0,8%	4300	4205	95	2,3%	620,1	592,2	27,9	4,7%
févr-89	162,2	161,3	0,9	0,6%	4059	4059	0	0,0%	595,3	591,5	3,8	0,6%
mars-89	185,8	176,7	9,1	5,1%	5058	4791	267	5,6%	678,9	661,6	17,3	2,6%
avr-89	197,6	202,5	-4,9	-2,4%	4300	4492	-192	-4,3%	728,6	700,2	28,4	4,1%
mai-89	204,2	201,2	3	1,5%	4870	4742	128	2,7%	720,8	727,1	-6,3	-0,9%
juin-89	206,8	206,7	0,1	0,0%	5559	5290	269	5,1%	827,5	775,7	51,8	6,7%
juil-89	249,3	250	-0,7	-0,3%	5190	5176	14	0,3%	786,4	732,1	54,3	7,4%
août-89	242,4	245,2	-2,8	-1,1%	4300	4426	-126	-2,8%	676,5	677,7	-1,2	-0,2%
sept-89	207,2	208,4	-1,2	-0,6%	4380	4396	-16	-0,4%	767	763,8	3,2	0,4%
oct-89	204,5	202,8	1,7	0,8%	4610	4470	140	3,1%	781,5	758,7	22,8	3,0%
nov-89	183,6	185,7	-2,1	-1,1%	4230	4206	24	0,6%	699	702,4	-3,4	-0,5%
déc-89	179,3	182,5	-3,2	-1,8%	4858	4928	-70	-1,4%	725,4	733,1	-7,7	-1,1%
janv-90	180,8	180,6	0,2	0,1%	4150	4291	-141	-3,3%	673,9	696,9	-23	-3,3%

Graphiques 5.1 à 5.5 : Réalisation, référence, et écart absolu.

Figure 5.1 : La circulation routière sur le réseau national, sur 1984-1991.
Unité : indice de débit base 100 en janvier 1970, Source : SETRA, snrd.

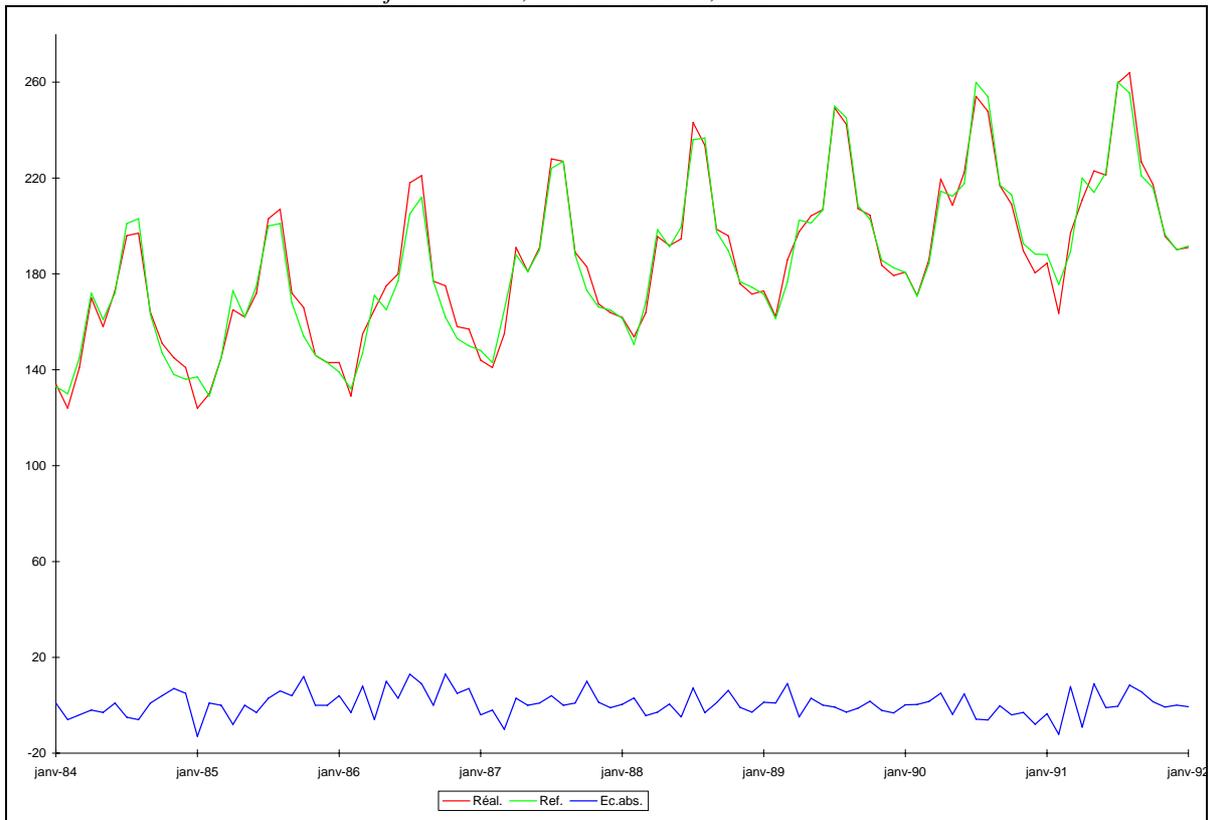


Figure 5.2 : Le trafic de voyageurs sur le réseau principal de la SNCF, sur 1984-1991.
Unité :million de voyageur-kilomètres, Source : SNCF.

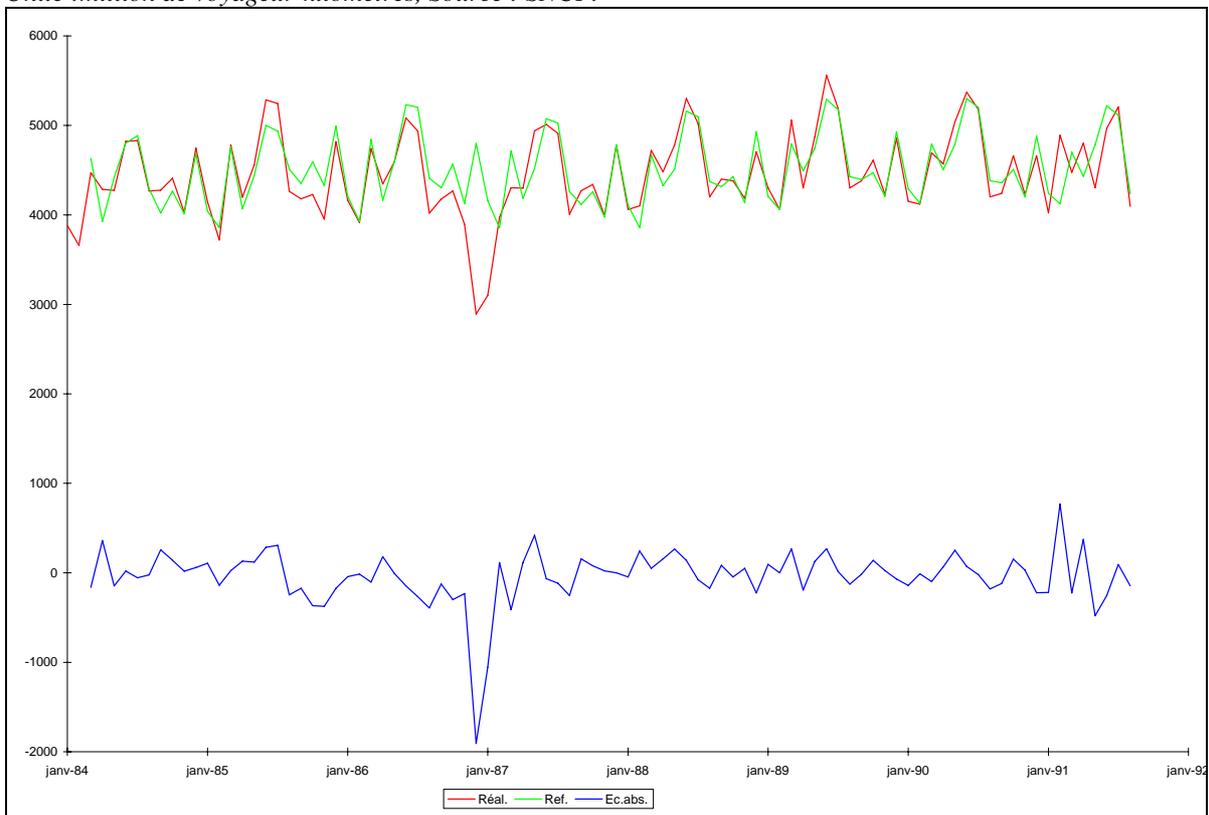


Figure 5.3 : Le trafic de passagers d'Air Inter, sur 1984-1991.

Unité : million de passager-kilomètres, Source : Air Inter.



Tableau 5.4: Les écarts Réalisation/Référence, absolus et relatifs, pour les deux indicateurs de trafics de fret (route, fer), sur 1986-1991

	Réal.	Réf.	Ec.abs.	Ec.rel.	Réal.	Réf.	Ec.abs.	Ec.rel.
juil-86	7,71				4,42			
août-86	4,94	4,94	0	0,0%	3,21	3,14	0,07	2,2%
sept-86	8,48	8,08	0,4	5,0%	4,42	4,45	-0,03	-0,7%
oct-86	9,3	8,98	0,32	3,6%	4,74	4,86	-0,12	-2,5%
nov-86	7,9	8,05	-0,15	-1,9%	4,27	4,64	-0,37	-8,0%
déc-86	7,9	6,96	0,94	13,5%	3,42	4,26	-0,84	-19,7%
janv-87	7,07	7,71	-0,64	-8,3%	3,38	4,22	-0,84	-19,9%
févr-87	7,48	7,23	0,25	3,5%	4,75	4,04	0,71	17,6%
mars-87	9,12	8,28	0,84	10,1%	4,75	4,61	0,14	3,0%
avr-87	9,1	8,54	0,56	6,6%	4,57	4,6	-0,03	-0,7%
mai-87	8	8,37	-0,37	-4,4%	4,05	4,31	-0,26	-6,0%
juin-87	9,6	8,59	1,01	11,8%	4,51	4,52	-0,01	-0,2%
juil-87	8,3	7,83	0,47	6,0%	4,21	4,2	0,01	0,2%
août-87	5,4	5,21	0,19	3,6%	3,26	3,04	0,22	7,2%
sept-87	9,1	8,89	0,21	2,4%	4,4	4,19	0,21	5,0%
oct-87	9,46	9,77	-0,31	-3,2%	4,85	4,63	0,22	4,8%
nov-87	9,28	8,58	0,7	8,2%	4,59	4,54	0,05	1,1%
déc-87	8,08	8	0,08	1,0%	4,3	4,3	0	0,0%
janv-88	8,01	8,61	-0,6	-7,0%	4,2	4,4	-0,2	-4,5%
févr-88	10,31	8,18	2,13	26,0%	4,28	4,4	-0,12	-2,7%
mars-88	11,07	9,2	1,87	20,3%	4,66	4,64	0,02	0,4%
avr-88	8,41	9,51	-1,1	-11,6%	4,4	4,52	-0,12	-2,7%
mai-88	9,49	9,15	0,34	3,7%	4,28	4,08	0,2	4,9%
juin-88	10,35	9,54	0,81	8,5%	4,55	4,52	0,03	0,7%
juil-88	9,32	8,78	0,54	6,2%	4,17	4,21	-0,04	-1,0%
août-88	6,8	5,94	0,86	14,5%	3,58	3,23	0,35	10,8%
sept-88	9,28	9,88	-0,6	-6,1%	4,56	4,5	0,06	1,3%
oct-88	10,07	10,62	-0,55	-5,2%	4,37	4,92	-0,55	-11,2%
nov-88	9,86	9,75	0,11	1,1%	4,72	4,59	0,13	2,8%
déc-88	8,66	8,67	-0,01	-0,1%	4,57	4,3	0,27	6,3%
janv-89	9,77	9,25	0,52	5,6%	4,62	4,28	0,34	7,9%
févr-89	9,48	9,29	0,19	2,0%	4,48	4,45	0,03	0,7%
mars-89	10,43	10,58	-0,15	-1,4%	4,88	4,82	0,06	1,2%
avr-89	10,92	10,92	0	0,0%	4,7	4,61	0,09	2,0%
mai-89	10,18	9,88	0,3	3,0%	4,45	4,39	0,06	1,4%
juin-89	10,8	10,56	0,24	2,3%	4,68	4,73	-0,05	-1,1%
juil-89	9,2	9,65	-0,45	-4,7%	4,25	4,37	-0,12	-2,7%
août-89	6,96	7,22	-0,26	-3,6%	3,58	3,34	0,24	7,2%
sept-89	9,98	10,35	-0,37	-3,6%	4,36	4,51	-0,15	-3,3%
oct-89	11,38	11,15	0,23	2,1%	4,62	4,88	-0,26	-5,3%
nov-89	10,96	10,32	0,64	6,2%	4,45	4,78	-0,33	-6,9%
déc-89	8,77	9,44	-0,67	-7,1%	4,15	4,46	-0,31	-7,0%
janv-90	10,41	10,03	0,38	3,8%	4,34	4,45	-0,11	-2,5%
févr-90	9,66	9,82	-0,16	-1,6%	4,26	4,45	-0,19	-4,3%
mars-90	11,13	10,53	0,6	5,7%	4,37	4,63	-0,26	-5,6%
avr-90	9,96	10,39	-0,43	-4,1%	4,15	4,48	-0,33	-7,4%
mai-90	10,55	10,01	0,54	5,4%	4,51	4,27	0,24	5,6%
juin-90	10,69	10,54	0,15	1,4%	4,63	4,7	-0,07	-1,5%
juil-90	9,89	9,84	0,05	0,5%	4,4	4,2	0,2	4,8%
août-90	6,77	7,12	-0,35	-4,9%	3,44	3,19	0,25	7,8%
sept-90	9,78	10,4	-0,62	-6,0%	4,06	4,44	-0,38	-8,6%
oct-90	10,96	11,14	-0,18	-1,6%	4,48	4,8	-0,32	-6,7%
nov-90	10,19	10,21	-0,02	-0,2%	4,47	4,61	-0,14	-3,0%
déc-90	8,33	8,96	-0,63	-7,0%	3,75	4,22	-0,47	-11,1%
janv-91	10,3	9,9	0,4	4,0%	4,58	4,28	0,3	7,0%
févr-91	9,61	9,6	0,01	0,1%	4,08	4,27	-0,19	-4,4%
mars-91	10,94	10,82	0,12	1,1%	4,51	4,31	0,2	4,6%
avr-91	10,9	10,03	0,87	8,7%	4,66	4,17	0,49	11,8%
mai-91	10,27	10,08	0,19	1,9%	4,16	4,09	0,07	1,7%
juin-91	10,63	10,75	-0,12	-1,1%	4,33	4,2	0,13	3,1%
juil-91	10,52	9,88	0,64	6,5%	4,21	3,91	0,3	7,7%
août-91	6,38	7,17	-0,79	-11,0%	3,53	3,07	0,46	15,0%
sept-91	10,55	10,28	0,27	2,6%	4,19	4,08	0,11	2,7%
oct-91	11,9	11,54	0,36	3,1%	4,85	4,46	0,39	8,7%
nov-91	10,59	10,8	-0,21	-1,9%	4,44	4,5	-0,06	-1,3%
déc-91	9,71	8,68	1,03	11,9%	4,4	4,04	0,36	8,9%

Figure 5.4: Le trafic routier de marchandises, sur 1986-1991.

Unité : milliard de tonnes-kilomètres, Source : OEST (enquête du transport routier de marchandises).

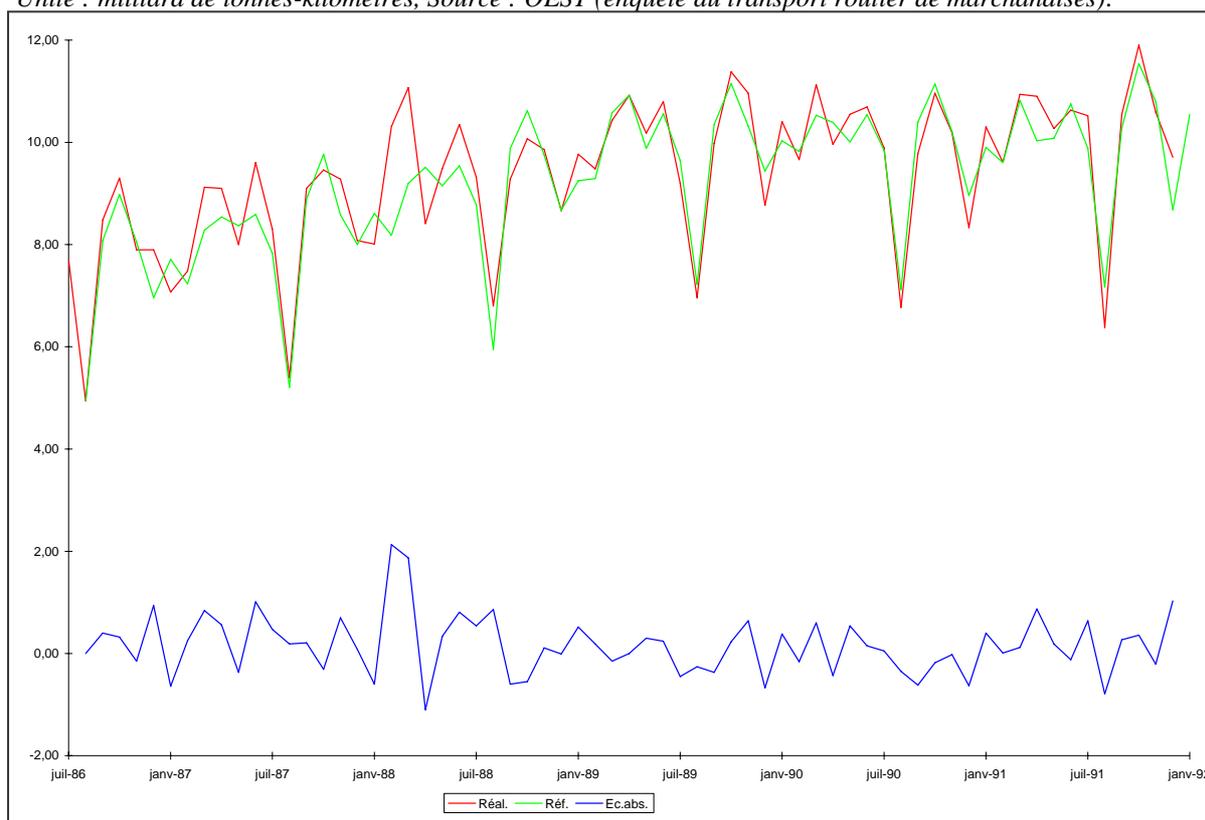


Figure 5.5 : Le trafic fret de la SNCF, sur 1986-1991.

Unité : milliard de tonnes-kilomètres, Source : SNCF



Chapitre 6.

Les effets de nature transitoire

Nous exposons ici, en les développant, les travaux présentés dans le cadre d'une communication à la 6^{ème} Conférence Mondiale sur la Recherche dans les Transports (Bergel, 1992).

A la différence de ce qui vient d'être exposé dans le chapitre 5 (appel à des variables indicatrices), les effets transitoires de nature climatique et calendaire que nous abordons dans ce chapitre seront traités de manière plus systématique (appel à des variables indicatrices et à des variables continues).

6.1. Objectif

L'on vise ici à affiner, en utilisant des modèles de séries temporelles avec variables exogènes, les effets transitoires dus à la configuration du calendrier ou à des conditions climatiques qui s'écartent de la normale saisonnière. Bien que transitoires, ces effets sont de fait importants pour l'analyse de l'évolution des flux de trafics.

6.2. Méthode

La question qui se pose est de modéliser simultanément ou séparément les effets du climat et du calendrier - ou encore « effet météorologique » et « effet calendaire » - sur les flux de trafic. Les variables exogènes utilisées pour représenter ces phénomènes transitoires, supposées connues, sont non corrélées. L'hypothèse faite est que leurs effets dans la modélisation sont additifs, de sorte que l'on peut estimer les deux types d'effets sur une variable de trafic de manière indépendante. Cette hypothèse se valide a posteriori.

La démarche séquentielle autorisée par l'indépendance a été retenue pour la modélisation journalière car elle est la seule possible en raison du grand nombre de données au niveau du jour. Elle permet en outre d'utiliser des techniques statistiques différentes pour chaque séquence, comme nous le verrons plus loin.

La démarche globale a quant à elle été menée sur une base mensuelle.

6.2.1. Configuration calendaire

Nous avons tenté de répondre aux questions suivantes :

- quel type d'effet modéliser ?
- avec quelles variables ?
- et comment les construire ?

On distingue :

- l'effet *périodique* qui se reproduit généralement de semaine en semaine ou d'année en année,
- l'effet *exceptionnel* non périodique qui se reproduit chaque année mais dans une configuration calendaire variable selon qu'il s'agit du calendrier des jours fériés ou du calendrier des congés (scolaires ou en entreprises).

C'est cet effet de calendrier exceptionnel sur lequel l'effort a porté.

La difficulté tient à la définition d'une classification des jours de l'année : une hypothèse doit être posée, qui sera validée par modélisation. La construction des indicatrices, qui constitue l'étape suivante, ne présente pas de difficulté en soi³¹.

Pour le transport de fret, il n'est pas apparu nécessaire de prendre en compte le calendrier des congés, mais seulement le calendrier des jours fériés. On retient pour ce faire 6 classes de jours : le jour ouvrable (du lundi au vendredi), le samedi et le dimanche, et trois types de jours fériés : le jour férié de semaine, le samedi férié et le jour pont. Sous une hypothèse de stabilité de structure, cette classification peut être exploitée directement dans un modèle mensuel, sans modélisation journalière préliminaire. On estime des poids journaliers additifs qui représentent l'effet additionnel du jour d'une classe par rapport à l'effet additionnel d'un jour ordinaire (cf. le Tableau 6.1). La somme sur un mois de tous ces poids journaliers mesure l'effet calendaire exceptionnel des jours du mois.

³¹ En pratique, et de manière générale, les effets de calendrier sont modélisés par l'adjonction dans le modèle ARIMA de variables indicatrices codant une typologie de jours de l'année. Par exemple, pour une série journalière, on introduit une variable égale à 1 le dimanche, et égale à 0 les autres jours, alors que pour une série mensuelle, on introduira une variable qui donne le nombre de dimanche du mois.

Pour le transport de voyageurs, la modélisation est plus fine.

Les effets calendaires pris en compte sont de plusieurs natures : un effet exceptionnel qu'on observe pendant les vacances scolaires, aux départs et aux retours de congés, au voisinage d'un jour férié, et un effet hebdomadaire. On modélise l'effet calendaire exceptionnel avec des indicatrices qui codent la présence d'une période de vacances scolaires, d'un jour de départ ou de retour de congés, ou la proximité d'un jour férié. Des hypothèses sur le nombre de jours affectés par la proximité d'un férié ont été retenues : 4 à 6 au voisinage du jour férié selon le cas. Ces hypothèses dépendent de la position du jour férié (cinq cas possibles en semaine), avec deux exceptions pour Pâques et la Pentecôte d'une part, pour l'Ascension d'autre part, qui se situent chaque année le même jour de semaine, et qu'on peut donc isoler pour affiner l'interprétation des coefficients calendaires (Azencott et al, 1992).

6.2.2. Météorologie

La question de la mesure du climat et de son effet n'est pas simple.

Nous avons tenté de répondre aux questions suivantes :

- quel type d'effet modéliser ?
- avec quels phénomènes météorologiques ?
- et avec quelles variables pour les mesurer ?

Sont à distinguer :

- l'effet régulier, saisonnier - *périodique*,
- et l'effet irrégulier, complémentaire du précédent, *non périodique*.

L'effet régulier est pris en compte dans la composante saisonnière, et représente un effet de normale saisonnière.

Nous nous intéressons ici à l'effet irrégulier, mesuré par différence avec l'effet de normale saisonnière. En pratique, la partie de normale saisonnière est éliminée par un filtre, ou estimée dans un prétraitement.

La difficulté tient ensuite au choix des phénomènes climatiques à retenir, puis à leur mesure³².

Il est important de noter que les variables météorologiques peuvent être construites à

³² En pratique, et de manière générale, les effets météorologiques sont aussi modélisés par l'adjonction de variables indicatrices - météorologie mesurée de manière qualitative -, ou directement par l'adjonction de variables continues - météorologie mesurée de manière quantitative. Par exemple, pour une série journalière, on introduit une variable égale à 1 chaque jour de gel, et égale à 0 les autres jours de l'année, et pour une série

l'extérieur du modèle, comme nous l'exposerons dans le chapitre 9 consacré au modèle Giboulée, dans lequel un exemple de modélisation journalière de l'effet climatique est donné.

Les variables météorologiques retenues dans la modélisation mensuelle sont celles qui sont apparues significatives parmi la dizaine de variables descriptives d'une situation climatique³³ fournies par Météo-France. Elles décrivent la température du jour (la température, fortement liée au kilométrage parcouru, rend bien compte de l'évolution de la mobilité à l'intérieur de l'année), l'occurrence de neige ou de gel (la circulation routière peut être sensiblement réduite pendant les périodes de froid prolongé), et la hauteur ou l'occurrence - absence/présence - de pluie (la pluie réduit la mobilité ou la reporte).

Enfin, la mesure initiale de ces variables quantitatives est réalisée par poste d'observation météorologique - une centaine de stations réparties sur le territoire, et par tranche de 6 heures au moins. Il est dans tous les cas nécessaire de déterminer le niveau d'agrégation spatio-temporel auquel cette information sera résumée.

On peut la moyenniser, par département, par région administrative ou par zone géographique définie selon un critère d'homogénéité climatique - un découpage de la France en 6 régions homogènes au plan climatique a déjà été réalisé, avec un choix de 6 stations pilotes représentatives de chacune des régions (Azencott et al., 1991). On peut aussi utiliser des moyennes nationales, obtenues par agrégation des données fournies par la centaine de stations. Bien que la seconde option soit limitative, elle est plus interprétable, et peut être retenue dans une première approche. Disposer à la fois de variables régionales et de moyennes nationales est probablement la meilleure formule, sous réserve que les mêmes variables soient mesurées sous ces deux formes, et utilisées ensemble pour la modélisation. C'est l'option qui a été retenue dans la version finale du modèle Giboulée opérationnel au SETRA, présenté dans le chapitre 9.

mensuelle on introduira une variable qui donne le nombre de jours de gel du mois. Dans le cas d'une variable continue, on utilisera pour une série journalière la température moyenne du jour, ou la hauteur de pluie du jour, et pour série mensuelle la température moyenne du mois et la hauteur de pluie moyenne du mois.

³³ Les variables climatiques utilisées, quantitatives ou qualitatives, sont ici associées à des phénomènes climatiques. Mais elles auraient pu être qualitatives, associées à des types de temps résultant d'une typologie très fine : cette approche qualitative de la météorologie en accidentologie routière (Guilain, 1993) n'a pas été retenue dans la version finale du modèle Giboulée.

6.3. Applications

Les deux applications que nous présentons portent d'une part sur les transports de marchandises, qui subissent l'effet du calendrier des jours fériés, avec quelques chocs isolés occasionnés par une grève à la SNCF, et par des périodes de froid prolongé pour le transport routier, et d'autre part sur les transports de voyageurs à travers la circulation des véhicules sur le réseau national, qui dépend plus particulièrement des conditions météorologiques, et qui subit l'effet du calendrier des jours fériés et du calendrier scolaire.

6.3.1. Trafic de fret

On dispose de statistiques mensuelles de trafic routier, réalisé par les véhicules français de 3 tonnes de charge utile, et de trafic de fret SNCF, homogènes sur longue période. Le rail et la route se répartissent l'essentiel des trafics terrestres de marchandises, soit 29% et 67% des trafics terrestres en 1991, 4% des trafics allant à la voie d'eau. Des modèles d'évolution mensuelle ont été établis sur 1975-1989, pour ces deux indicateurs, afin de dégager les caractéristiques propres à chacun des modes.

On modélise d'abord les effets de la grève de la SNCF de décembre 1986-janvier 1987 (d'une durée de quatre semaines, réparties sur les deux mois) pour le trafic de fret ferroviaire, puis les effets de cette grève et des périodes de froid prolongé pour le transport routier de marchandises, avec des variables d'intervention. L'effet du calendrier sur les deux indicateurs de trafic est modélisé sur la base d'hypothèses formulées par la SNCF, qui consistent à distinguer six classes de jour dans l'année. L'examen des statistiques hebdomadaires (les statistiques journalières ne sont pas disponibles) confirme l'effet déterminant des jours fériés, plus important en moyenne que celui d'une grève ordinaire.

Nous donnons à titre d'exemple le modèle établi pour le fret ferroviaire (le modèle relatif au fret routier est similaire), sur la période janvier 1975 - décembre 1989 :

$$(I - B)(I - B^{12}(I + \underset{(0.07)}{0.45}B + \underset{(0.08)}{0.42}B^2))(Y_t - Z_t) = (I - \underset{(0.07)}{0.52}B^{12})\varepsilon_t,$$

$$Z_t = \underset{(0.16)}{-0.93} - \underset{(0.17)}{0.91}B + \underset{(0.16)}{0.33}B^2)GREVE_{déc86-janv87} + \underset{(0.08)}{0.17}SA_t + \underset{(0.08)}{0.12}DI_t + \underset{(0.08)}{0.17}FE_t + \underset{(0.09)}{0.11}SF_t + \underset{(0.08)}{0.22}PO_t + \underset{(0.08)}{0.26}N_t$$

avec : Y_t la série mensuelle de trafic, mesurée en milliards de tonnes-kilomètres,

SA le nombre de samedis du mois,

DI le nombre de dimanche du mois,

Fe le nombre de jours fériés du mois,
 SF le nombre de samedis fériés du mois,
 PO le nombre de jours pont (lundi ou vendredi) du mois,
 N le nombre des autres jours du mois,
 et u_t le bruit blanc résiduel, d'écart-type $\sigma_u = 0,2$
 sachant que $Y_{\text{moy}} = 4,9$, à partir de 180 observations.

La perte de trafic due à la grève est estimée à 0,9 milliard de tonnes-kilomètres pour chacun des deux mois de décembre 86 et janvier 87, et suivie d'un report partiel de trafic estimé à 0,3 milliard de tonnes-kilomètres en février.

Le tableau 6.1 donne les coefficients calendaires obtenus pour les deux modèles relatifs au trafics de fret, pour le fer et la route, et les poids calculés par classe de jour. Les poids journaliers proposés pour le transport ferroviaire sont cohérents avec les ordres de grandeur proposés par la SNCF, bien que plus élevés. Les poids journaliers obtenus pour le transport routier sont encore plus élevés : on remarquera que les transports routiers effectués sur plusieurs jours peuvent être enregistrés au moment du départ, par exemple un dimanche soir (les interdictions de circuler valent jusqu'à 22 heures).

Tableau 6.1 : Poids journaliers pour les trafics de fret, estimés sur 1975-1989

<i>Effets et poids journaliers</i>	<i>Transport ferroviaire</i>		<i>Transport routier</i>	
	<i>effet</i>	<i>poids</i>	<i>effet</i>	<i>poids</i>
jour de semaine	0,26	1	0,62	1
samedi	0,17	0,65	0,54	0,86
dimanche	0,12	0,45	0,52	0,84
jour férié de semaine	0,17	0,66	0,44	0,71
samedi férié	0,11	0,42	0,49	0,80
jour pont(lundi ou vendredi)	0,22	0,85	0,51	0,83

Unité : milliard de tonnes-kilomètres

Source : SNCF et OEST (enquête du transport routier de marchandises)

6.3.2. Circulation routière

La circulation inter-urbaine est mesurée à partir de comptages des parcours effectués sur le réseau national (routes et autoroutes) tous véhicules confondus, toutefois les voitures particulières y occupent la plus large part - 70% environ, d'après le recensement 1990 du SETRA. On observe de ce fait une saisonnalité marquée – on circule, en moyenne, plus à partir de Pâques et jusqu'à la Toussaint, avec un niveau record en juillet et août - propre au

transport de voyageurs, qui s'explique à la fois par les variations climatiques à l'intérieur de l'année et par la configuration du calendrier. On dispose ici de statistiques journalières fiables, à partir de 1985, sur les routes nationales et les autoroutes concédées, de sorte qu'une modélisation a pu être réalisée sur six ans.

Deux approches ont été menées : d'une part une modélisation des séries journalières de parcours effectués sur les routes nationales et les autoroutes concédées (soit 54% et 29% des parcours effectués sur l'ensemble du réseau national en 1990), de janvier 1985 à décembre 1990, et d'autre part une modélisation directe des parcours mensuels effectués sur routes nationales, sur autoroutes concédées et sur l'ensemble du réseau, sur plus longue période 1972-1990.

6.3.2.1. Approche journalière

Sur une base journalière, le grand nombre de données disponibles sur les six ans aurait permis d'opter pour une approche globale, dans laquelle les effets de nature climatique et calendaire peuvent être estimés simultanément.

Mais une approche séquentielle a été retenue, qui est détaillée dans le chapitre 9. Cette méthode s'applique en effet à tous types d'indicateurs journaliers, et a été utilisée directement sur des données journalières de sécurité routière, dans le modèle Giboulée qui consiste à désaisonnaliser les statistiques journalières d'accidents corporels et de victimes de la circulation routière, et qui est exploité mensuellement par le SETRA.

La méthode consiste à réaliser un prétraitement local des données à modéliser en les corrigeant de manière fine de l'influence du climat et à estimer ensuite les effets calendaires de deux natures différentes : l'effet calendaire exceptionnel qui s'observe au voisinage des jours fériés et des périodes de congés (calendrier férié et calendrier scolaire), et l'effet de calendrier hebdomadaire qui s'observe de manière régulière à l'intérieur de la semaine.

Nous donnons ici la partie du modèle journalier qui porte sur l'estimation de l'effet calendaire exceptionnel, après le prétraitement des données qui a permis de les corriger de l'influence locale du climat.

Le modèle général suivant a été établi pour chacun des deux indicateurs de parcours, sur routes nationales et sur autoroutes concédées, pour la période de janvier 1985 à décembre 1990 (soit 2191 observations) :

$$(I - B^7)(I - aB)(YC_t - \sum_{i=1}^I a_i X_{it}) = (I - bB)(I - cB^7)u_t,$$

avec : YC_t la série des parcours corrigée de l'effet météorologique en t,

X_{it} , une variable calendaire en t, et a_i son coefficient, pour $i=1$ à I,

$\sum_{i=1}^I a_i X_{it}$ la somme des effets calendaires instantanés en t,

$YC_t - \sum_{i=1}^I a_i X_{it}$ la série corrigée des effets calendaires instantanés en t,

et u_t le bruit blanc résiduel

On dispose ainsi de « profils d'impact » du calendrier des jours fériés, et du calendrier scolaire, sur le trafic journalier sur les routes nationales et les autoroutes concédées (cf. la figure 6.1 qui donne les profils journaliers du 1^{er} septembre au 31 décembre 1990). Les coefficients calendaires estimés par ce modèle sont pratiquement tous très significatifs, (cf. le tableau 6.2). Alors que le trafic sur autoroutes concédées est de l'ordre de la moitié du trafic sur routes nationales sur la période, on notera que les suppléments de trafic enregistrés les jours de grands déplacements sont du même ordre sur les deux réseaux - ou encore que les suppléments de trafic sont, en variations relatives, deux fois plus importantes sur autoroutes concédées.

6.3.2.2. Approche mensuelle

Nous donnons ici le modèle retenu pour la circulation routière enregistrée sur les routes nationales (le modèle retenu pour la circulation enregistrée sur les autoroutes concédées est semblable), pour la période janvier 1983 à décembre 1990 :

$$(I - B)(I - B^{12})(Y_t - Z_t) = (I - 0,72B)(I - 0,21B^{12})u_t,$$

$$Z_t = 0,037TMAX_t - 0,1724NNEIGE_t + 1,59DHIV_t + 1,45DPRIN_t + 0,85DETE_t + 0,99X3_t + 0,2662N_t$$

avec : Y_t la série mensuelle des parcours sur routes nationales, mesurée en centaines de millions de véhicules-kilomètres,

, $TMAX$ la température maximale journalière, moyennée sur le territoire et sur le mois,

$NNEIGE$ l'occurrence de neige journalière, moyennée sur le territoire et sur le mois,

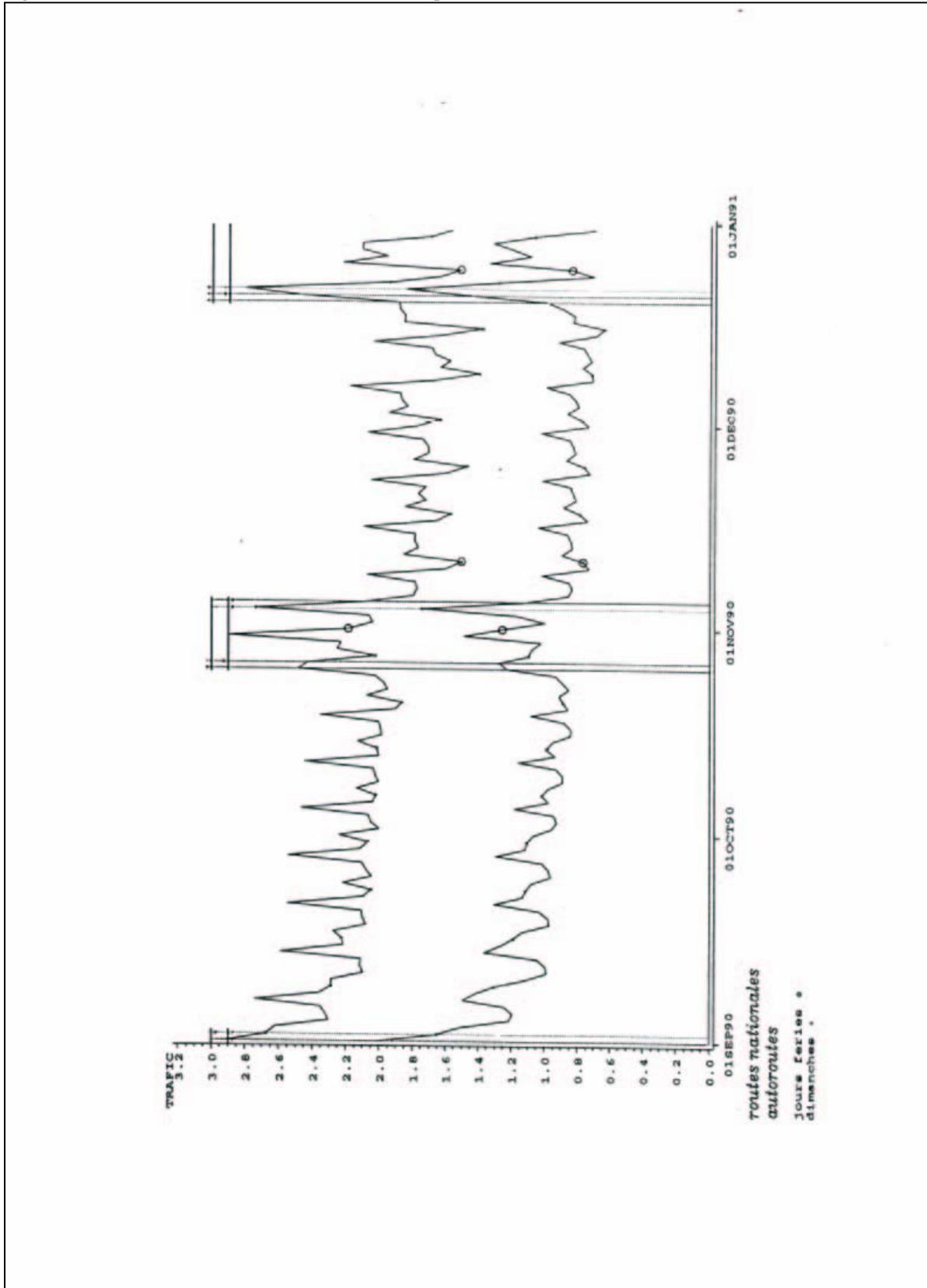
DHIV le nombre de jours de départs et de retours de vacances d'hiver dans le mois,
 DPRIN le nombre de jours de départs et de retours de vacances de Pâques dans le mois,
 DETE le nombre de jours de départs et de retours de vacances de pâques dans le mois,
 X3 le nombre de jours du mois affectés, de manière positive et importante, par la proximité d'un jour férié,
 N le nombre des autres jours du mois,
 et u_t le bruit blanc résiduel, d'écart-type $\sigma_u = 1,1$
 sachant que $Y_{\text{moy}} = 58$, à partir de 96 observations.

Les graphiques 6.2 et 6.3 suivants fournissent pour 1989 et 1990 des estimations des séries mensuelles de parcours, sur routes nationales et sur autoroutes concédées, issues de modèles sans variable externe (estimations modèle ARMA) et de modèles avec variables météorologiques et calendaires (estimations modèle ARMAX). Les profils des estimations issues des modèles avec variables externes se rapprochent très sensiblement des profils des données brutes. Pour l'année 1990, la circulation a été ralentie à l'été et en fin d'année (le trafic a été perturbé en décembre par des chutes de neige d'une ampleur exceptionnelle). Sur l'ensemble de l'année 1990, la croissance de la circulation sur l'ensemble du réseau national serait de 4,4% en données corrigées des effets météorologiques et calendaires, alors qu'elle est de 3,9% en données brutes. Ce redressement provient du réseau des routes nationales, pour lequel la croissance serait de 2,1% en données corrigées, alors qu'elle est de 1,6% en données brutes. Ces différentiels de croissance apparaissent relativement importants à un moment où le rythme de croissance de la circulation routière est sensiblement affecté par le ralentissement de l'activité économique induit par la crise du Golfe.

6.4. Discussion

Le premier résultat de cette approche avec variables externes, d'ordre statistique, est une amélioration de la qualité d'ajustement du modèle : pour les deux modélisations mensuelles, la variance résiduelle est sensiblement réduite - de 0,056 à 0,041 pour le modèle du trafic de fret ferroviaire (soit un écart-type résiduel réduit de 14,6%), et de 3,1 à 1,3 pour le modèle de la circulation routière sur routes nationales (soit un écart-type résiduel réduit de 35,2%). Il faut noter que l'estimation mensuelle, en moyenne meilleure avec le modèle amélioré, peut certains mois être très proche de la réalisation, de sorte que le profil mensuel de la série se trouve mieux reproduit, comme le montrent les graphiques 6.2 et 6.3 .

Figure 6.1 : Les relations trafic-calendrier, du 1^{er} septembre au 31 décembre 1990.



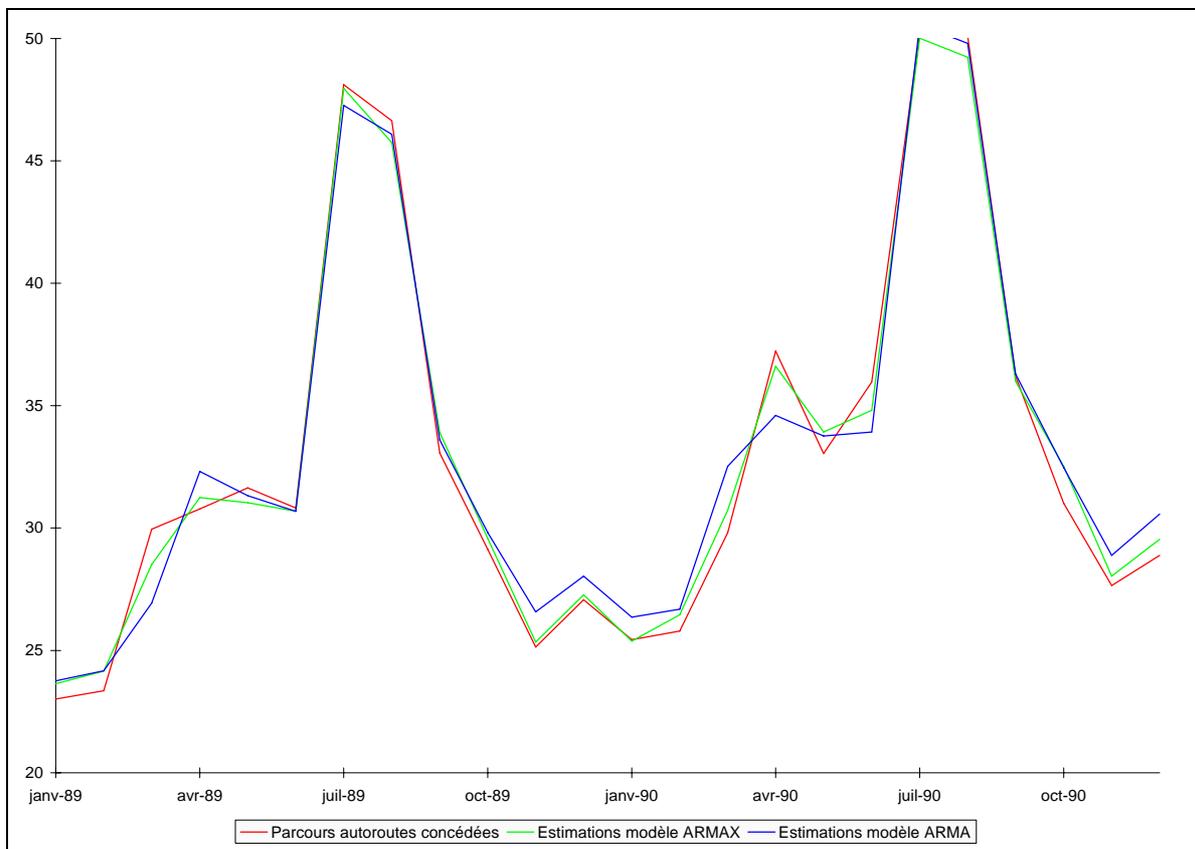
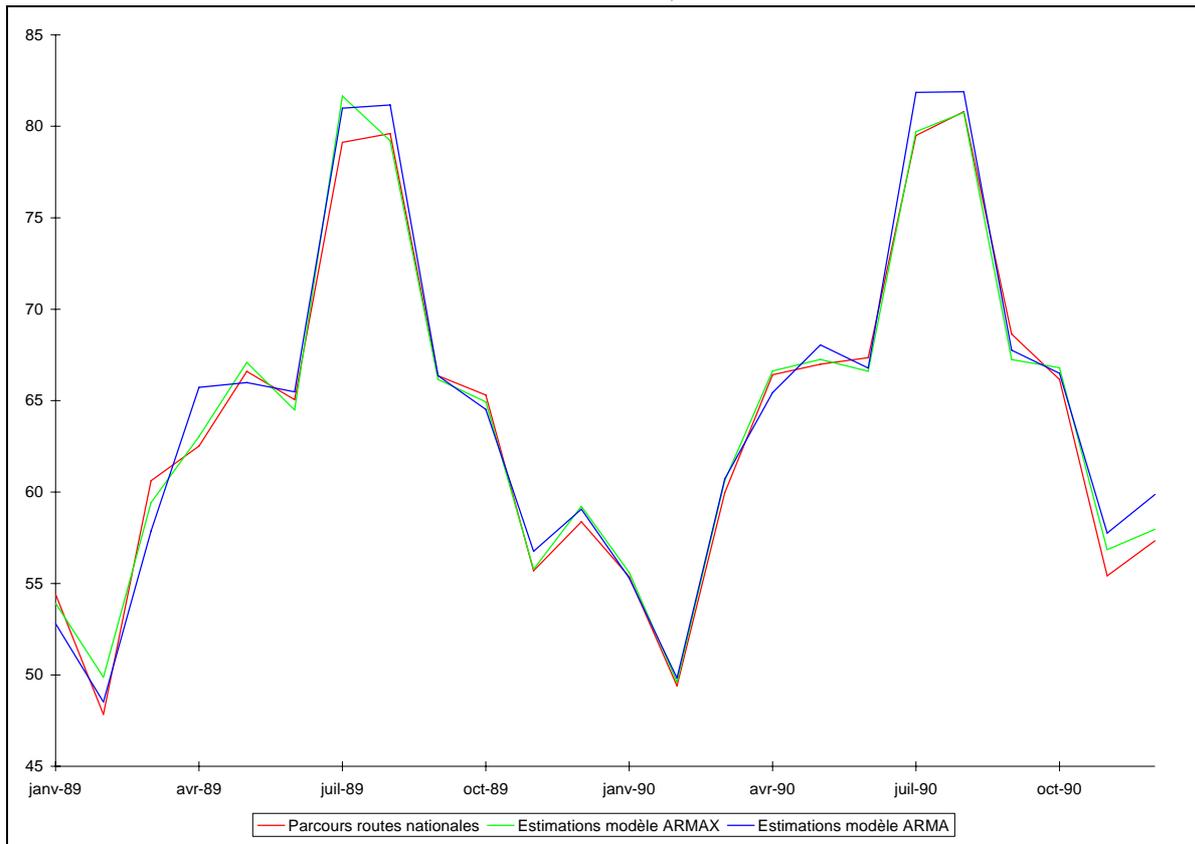
Source : Modélisation ARIMA d'une dizaine d'indicateurs d'insécurité routière. Rapport OEST/CEMS. Paris, OEST, 1992.

Tableau 6.2 : Coefficients calendaires pour la circulation routière interurbaine, estimés sur 1985-1990

Coefficients calendaires		Routes nationales	Autoroutes concédées	
Changements de niveau:				
Grandes vacances		0,05	0,06	
Petites vacances		0,15	0,16	
Noël		-0,55	-0,35	
Départs et retours de:				
printemps		0,15	0,10	
été		0,28	0,21	
Noël		0,22	0,22	
automne		0,24	0,14	
hiver		0,38	0,35	
Jours fériés:				
Pâques et Pentecôte	<i>vendredi</i>	0,44	0,35	
	<i>samedi</i>	0,61	0,37	
	<i>dimanche</i>	-0,20	-0,09	
	<i>lundi</i>	0,45	0,37	
	<i>mardi</i>	0,37	0,25	
Ascension	<i>mercredi</i>	0,62	0,40	
	<i>jeudi</i>	0,19	0,27	
	<i>vendredi</i>	-0,27	-0,22	
	<i>samedi</i>	-0,17	-0,15	
dimanche	<i>dimanche</i>	0,55	0,36	
	lundi férié	<i>vendredi</i>	0,27	0,12
		<i>samedi</i>	0,45	0,26
		<i>dimanche</i>	-0,23	-0,20
<i>lundi</i>		0,27	0,28	
mardi férié	<i>mardi</i>	0,27	0,24	
	<i>vendredi</i>	0,17	0,16	
	<i>samedi</i>	0,50	0,33	
	<i>dimanche</i>	0,00	0,02	
jeudi férié	<i>lundi</i>	-0,14	-0,13	
	<i>mardi</i>	0,26	0,27	
	<i>mercredi</i>	0,33	0,29	
	<i>jeudi</i>	0,20	0,07	
mercredi férié	<i>mercredi</i>	-0,11	0,01	
	<i>jeudi</i>	0,01	-0,05	
	<i>vendredi</i>	-0,12	-0,09	
	<i>mercredi</i>	0,48	0,27	
jeudi férié	<i>jeudi</i>	-0,03	0,06	
	<i>vendredi</i>	-0,17	-0,06	
	<i>samedi</i>	0,12	0,17	
	<i>dimanche</i>	0,51	0,33	
vendredi férié	<i>jeudi</i>	0,53	0,29	
	<i>vendredi</i>	-0,27	-0,08	
	<i>samedi</i>	-0,07	0,07	
	<i>dimanche</i>	0,50	0,33	
lundi	<i>lundi</i>	0,19	0,12	

Unité : 100 millions de véhicules-kilomètres
Données SETRA, snrd.

Figure 6.2 et Figure 6.3 : La circulation routière sur routes nationales et sur autoroutes concédées, sur 1989-1990
Unité : 100 millions de véhicules-kilomètres. Données SETRA, snrd.



Le second résultat est la quantification de la relation entre les variables exogènes et la variable modélisée, qu'il s'agisse d'une variable d'intervention, d'une variable calendaire ou encore d'une variable météorologique. La relation peut être difficile à interpréter, comme l'influence du climat sur la circulation routière à partir d'un modèle mensuel : on connaît par exemple, comme nous l'avons vu en 6.3.2.2, l'effet d'un degré supplémentaire de la température journalière maximale, moyennée sur le territoire et sur le mois, mais cette variation recouvre des évolutions locales forcément différentes. Cette limite a été contournée par l'approche journalière, dans laquelle la météorologie est mesurée localement, cette approche est exposée dans le chapitre 9.

On dispose enfin de l'effet global des variables externes, mesuré journallement ou mensuellement sur la période considérée. On peut ainsi mesurer séparément les effets de certains groupes de variables, en supposant ces effets indépendants.

Chapitre 7.

Les effets à court terme des déterminants de la demande

Nous présentons ici des travaux réalisés pour modéliser les effets infra-annuels des déterminants de la demande de transport, représentatifs de l'offre (prix, consistance des réseaux) ou de la demande (production/consommation) de transport sur les flux de trafics de voyageurs et de marchandises, agrégés par mode de transport et/ou par catégorie de réseau.

Les références reprises dans ce chapitre sont tirées de notes de Synthèse OEST-SES-SESP (SESP, 2008a), du rapport du groupe de travail Beauvais Consultants/DIAM-Recherche/DTT/DAEI/DRAST synthétisant des travaux de recherche financés par la DAEI et par la DRAST pour les Directions du Ministère de l'Équipement (Azencott et al, 1997a et b), et enfin de deux rapports de convention Inrets/DAEI réactualisant en 2000 et 2002 les travaux produits en 1997 (Bergel, Mutter, 2000 et Bergel, 2002). Une présentation synthétique des résultats obtenus en est faite dans (Bergel, 2003).

7.1. Problématique

Comme indiqué dans le chapitre 4, l'approche retenue depuis les années 1970 au Ministère de l'Équipement pour quantifier les relations entre les principaux déterminants de la demande de transport et les indicateurs de demande est menée sur une base annuelle, avec des modèles économétriques de régression linéaire multiple - ou s'y ramenant moyennant une transformation préliminaire des données utilisées. On peut ainsi réaliser des scénarios d'évolution des flux de trafic, à un horizon variant entre une et plusieurs années, sous réserve de disposer de scénarios d'évolution de leurs déterminants à ce même horizon.

Des prévisions de l'évolution annuelle des trafics de voyageurs et de marchandises, agrégés par mode de transport et par catégorie de réseau, sont systématiquement établies pour l'année en cours et pour l'année à venir. Les mêmes types de modèles peuvent être utilisés pour les prévisions à 1-2 ans et pour les prévisions à plus long terme, sur la base d'hypothèses relatives à la croissance économique, d'une part, et aux évolutions des caractéristiques du secteur des transports - consistance des réseaux et prix - (SESP, 2008a).

Pour autant, les déterminants de la demande de transport à long terme peuvent aussi se manifester sur les flux de trafics à un horizon inférieur à l'année. De fait, une augmentation importante du prix des carburants, des tarifs ferroviaire et aérien, une modification du réseau (extension du réseau à grande vitesse), etc. peuvent se traduire par une modification immédiate des niveaux attendus de trafic et par des reports modaux. Ces modifications peuvent ne se stabiliser qu'à terme.

L'analyse de ces effets de court terme a été menée en intégrant dans des modèles à base mensuelle/trimestrielle les déterminants de la demande à long terme. Le concept d'une modélisation multi-modale des trafics (route/fer/air pour les déplacements de voyageurs d'une part, route/fer/voies navigables pour les trafics de marchandises) qui s'imposait a priori conduisait à se doter de bases de données mensuelles ou trimestrielles, exhaustives et homogènes entre les modes de transport. Cette approche n'a pas été possible pour le mode aérien³⁴, qui n'a donc pas été retenu dans ces travaux. La voie d'eau n'a pas été retenue en raison de sa moindre part modale du trafic de fret.

La comparaison entre modèles annuels et modèles infra-annuels, utilisés sur les mêmes données mesurées à des échelles de temps différentes, a permis de mieux comprendre le mode de propagation des effets durables.

7.2. Objectif

Le premier objectif consiste à prendre en compte, sur une base infra-annuelle, les principaux déterminants représentatifs de l'offre (prix, consistance des réseaux) et de la demande (production/consommation) de transport, dans un modèle explicatif des trafics agrégés de voyageurs et de marchandises.

Le second objectif consiste à tirer des enseignements de la comparaison des résultats obtenus avec ceux des modèles annuels.

L'on se limite ici aux deux modes routier et ferroviaire, pour les transports interurbains de voyageurs et pour les trafics de marchandises. La période initiale de calibrage des modèles est

³⁴ L'analyse de l'évolution du trafic aérien domestique a néanmoins été menée sur une base mensuelle par une approche différente, visant à intégrer les effets de la concurrence modale (Pinon et al, 1995)(Bergel et al, 1995).

1980-1992³⁵, sur une base mensuelle et trimestrielle. Une actualisation des modèles sur la période 1980-1999 a été réalisée, sur une base trimestrielle et semestrielle, et pour un petit nombre d'indicateurs de trafic.

7.3. Méthode

Les transports interurbains de voyageurs sont mesurés par cinq indicateurs : d'une part la circulation des véhicules légers et des poids lourds, mesurés indistinctement sur routes nationales, sur autoroutes concédées et sur l'ensemble du réseau routier national (mesurés à partir des boucles magnétiques installées sur routes nationales et sur autoroutes), d'autre part la circulation des seuls véhicules légers sur autoroutes concédées (mesurée à partir des mesures de péage), et enfin le trafic de voyageurs sur le réseau principal de la SNCF.

On modélise trois types d'effets: *un effet de croissance économique* (mesuré par le produit intérieur brut pour la circulation routière, et par la consommation des ménages pour le transport ferroviaire), *un effet de réseau* (mesuré par la longueur du réseau autoroutier concédé, et par l'offre des TGV), et *un effet de prix* (mesuré par le prix des carburants voiture, le péage autoroutier et la recette ferroviaire au voyageur-kilomètre).

Les trafics globaux de marchandises sont mesurés par quatre indicateurs : les trafics de fret routier et ferroviaire sur le champ intérieur³⁶, et sur l'ensemble des deux champs intérieur et international.

On modélise deux types d'effets: *un effet de croissance économique* (mesuré par la production industrielle, décomposée en quatre secteurs distincts) et *un effet de prix* (prix du transport routier de marchandises en zone longue, et recette ferroviaire à la tonne-kilomètre transportée pour le trafic wagons).

7.3.1. Spécification

Afin de disposer d'une base de comparaison, des modèles ARIMA avec variables exogènes sont d'abord estimés dans une approche modale pour chacun des deux modes routier et ferroviaire ; puis des modèles vectoriels autorégressifs avec variables exogènes sont estimés dans une approche bimodale route/fer. La spécification est toujours linéaire, de sorte que les

³⁵ les statistiques des sociétés d'autoroutes à péage ne sont pas fournies avant 1980 ; et en raison de la récession économique et de la modification du système de distribution ferroviaire du début 1993, la période de calage a été limitée à 1980-1992.

³⁶ Intérieur ou domestique

élasticités du trafic au facteur exogène de demande sont mesurés par des paramètres constants, et donc par des élasticités variables avec le temps.

7.3.1.1. Cas univarié

La forme ARMAX suivante est estimée :

$$\Phi(B) \left[Y_t - \sum_{i=1}^K \Phi_i(B) Z_{it} \right] = \Theta(B) W_t$$

avec: Y la variable endogène à modéliser (éventuellement filtrée par un filtre de différence F(B)),

Z_i les K variables exogènes (éventuellement filtrées par des filtres F_i(B)),

W un bruit blanc non corrélé avec le passé de Y et des Z_i,

et Φ , Φ_i , Θ des polynômes retards en B.

7.3.1.2. Cas vectoriel

Dans le cas vectoriel, où le vecteur endogène Y a 2 composantes, on utilise la forme structurelle suivante :

$$\Phi(B) Y_t = \Psi(B) Z_t + \Theta(B) W_t$$

avec: Y et W les vecteurs endogène et de bruit blanc à 2 composantes,

Z le vecteur exogène à K composantes,

et Φ , Ψ , Θ des matrices dont chaque terme est un polynôme en B

ou encore une autre forme réduite, qui se déduit de la précédente, qui peut être préférée pour des raisons de commodité algorithmique mais au détriment du nombre total de paramètres du modèle :

$$\Theta^{-1} \Phi(B) Y_t = \Theta^{-1} \Psi(B) Z_t + W_t$$

c'est-à-dire : $\Phi'(B) Y_t = \Psi'(B) Z_t + W_t$

sous l'hypothèse que les matrices de polynômes Φ et Ψ se factorisent en produit de matrices de polynômes : $\Phi = \Theta\Phi'$ et $\Psi = \Theta\Psi'$.

Nous utilisons ici une méthode et un algorithme qui permettent de déterminer de façon automatique les ordres des matrices de polynômes autorégressifs et moyennes mobiles Φ et Θ , à partir du logiciel MODEST (Modélisation de la Dynamique d'un Ensemble de Séries Temporelles) (Azencott et al, 1997c).

7.3.2. Comparaison avec les modèles annuels

La cohérence de ces modèles avec des modèles économétriques similaires à périodicité annuelle a été testée de la manière suivante : des élasticités annuelles sont calculées par simulation de l'évolution des trafics à un horizon d'une année, et comparées avec les élasticités annuelles obtenues avec d'autres modèles à pas annuel calés sur les mêmes données.

7.3.3. Validation statistique et performance empirique

La validation statistique est réalisée à partir des tests habituels de significativité des paramètres, et d'hypothèse sur le résidu du modèle, donnés en 2.1.1.4.

La performance empirique est évaluée à partir de critères, tels que le R^2 ³⁷ pour le pouvoir d'ajustement, et en particulier le MAPRE pour le prévoir de prédiction, donnés en 2.1.1.4. également.

Un critère de performance empirique à un an - c'est-à-dire à 4 pas sur une année civile - a été considéré (en complément du critère performance à un pas, et ce toujours pour des raisons de comparaison avec la performance des modèles annuels établis sur les mêmes données. Dans les applications qui suivent, des critères de performance empirique à 1 an sont donnés afin de comparer la performance de modèles univarié et bivarié utilisés sur les mêmes données (cf. le tableau 7.6 pour les trafics intérieurs de marchandises, transportés par route et par fer). La comparaison avec la performance des modèles annuels utilisés sur les mêmes données de trafic, mesurés en rythme annuel, n'est pas mentionnée dans ce chapitre, mais elle est a été effectuée et a permis de conclure à la supériorité des modèles établis sur données trimestrielles.

³⁷ L'usage du R^2 est plus fréquent dans les modèles de régression que dans les modèles autorégressifs. Il lui est aujourd'hui préféré celui d'un critère d'information pénalisant.

7.4. Applications

Des résultats portant sur l'ensemble des indicateurs modélisés, définis en 7.3, sur une base mensuelle, trimestrielle et annuelle sont donnés dans le rapport du groupe de travail (Azencott et al, 1997a et b). Nous nous limitons ici à donner deux exemples d'applications établies sur une base trimestrielle et sur 1980-1992, pour les transports intérieurs de fret puis pour les transports interurbains de voyageurs.

7.4.1. Trafics intérieurs de marchandises

Le premier exemple porte sur les trafics intérieurs de marchandises, pour la route et pour le fer, sur une base trimestrielle.

7.4.1.1. Les variables

Une base de données mensuelles a été constituée, dont a été tirée une base trimestrielle - par somme ou par moyenne des données mensuelles - sur la période 1980-1992.

<i>Trm</i>	Trafic intérieur routier de marchandises, mesuré en millions de tonnes-km, source SES (enquête du TRM)
<i>Sncf</i>	Trafic intérieur ferroviaire de marchandises, mesuré en millions de tonnes-km, source SNCF
<i>Ipl</i>	Indice de production des IAA, base 100 en 1985, source INSEE
<i>ip3</i>	Indice de production des biens intermédiaires, base 100 en 1985, source INSEE
<i>ip7</i>	Indice de production du bâtiment, génie civil et agricole, base 100 en 1985, source INSEE
<i>ip9</i>	Indice de l'ensemble de l'industrie hors bâtiment et génie civil et agricole, base 100 en 1985, source INSEE
<i>ip93</i>	Indice de l'ensemble de l'industrie hors bâtiment et génie civil et agricole et hors biens intermédiaires, base 100 en 1985, source INSEE
<i>ip931</i>	Indice de l'ensemble de l'industrie hors bâtiment et génie civil et agricole et hors biens intermédiaires et hors IAA, base 100 en 1985, source INSEE,
<i>Iprout</i>	Prix du transport routier de marchandises zone longue, TRO avant 1985 puis enquête TRM, en francs constants de 1980, source DTT/SES
<i>Iprmtk</i>	Recette unitaire WAGONS, en francs constants de 1980 à la tonne-km, source SNCF

Tableau 7.1 - Variables utilisées dans l'exemple des modèles de trafics de marchandises

7.4.1.2. Modèle univarié du trafic routier de fret

$$(I - B^4)(I - \underset{(0.16)}{0.33}B - \underset{(0.17)}{0.26}B^2)(trm_t - Z_t) = (I - \underset{(0.17)}{0.54}B^4)\varepsilon_t - 290.6$$

$$\text{avec les corrections } Z_t = \underset{(59.95)}{116.25}ip3 + \underset{(32.97)}{77.13}ip7_t + \underset{(99)}{95.5}ip93_t - \underset{(7592.8)}{21832.8}iprout_t$$

écart-type résiduel=662.4

part de variance expliquée=76.8%

	1984	1992
<i>ip3</i>	0.52	0.41
<i>ip7</i>	0.42	0.39
<i>ip93</i>	0.37	0.32
<i>iprout</i>	-0.85	-0.46

Tableau 7.2 - Elasticités apparentes pour le modèle univarié du trafic routier de fret, sur 1980-1992

7.4.1.3. Modèle univarié du trafic ferroviaire de fret

Les variables *ip7* et *iprout* s'excluent mutuellement. On a donc 2 modèles alternatifs :

(a) Avec le prix du mode routier :

$$(I - B^4)(I + 0.60B^2)(sncf_t - Z_t) = (I - 0.83B^4)\varepsilon_t - 318.3$$

$$\text{avec } Z_t = -33.3ip1_t + 50.3ip3_t + 48.2ip931_t - 35633.4iprmtk_t + 3924.7iprout_t$$

écart-type résiduel=241.7

part de variance expliquée=58.5%

	1984	1992
<i>ip1</i>	-0.38	-0.51
<i>ip3</i>	0.58	0.73
<i>ip931</i>	0.55	0.74
<i>Iprmtk</i>	-0.66	-0.60
<i>Iprout</i>	0.32	0.28

Tableau 7.3 - Elasticités apparentes pour le modèle univarié du trafic ferroviaire de fret, sur 1980-1992 (a)

(b) Avec la variable de production du BTP :

$$(I - B^4)(I - 0.15B + 0.48B^2)(sncf_t - Z_t) = (I - 0.78B^4)\varepsilon_t - 370$$

$$\text{avec } Z_t = -15ip1_t + 43.9ip3_t - 11.2ip7_t + 44ip931_t - 28683.5iprmtk_t$$

écart-type résiduel=249.4

part de variance expliquée=55.8%

	1984	1992
<i>ip1</i>	-0.17	-0.23
<i>ip3</i>	0.51	0.64
<i>ip7</i>	-0.13	-0.19

<i>ip931</i>	0.50	0.67
<i>iprmtk</i>	-0.53	-0.49

Tableau 7.4 - Elasticités apparentes pour le modèle univarié du trafic ferroviaire de fret, sur 1980-1992 (b)

7.4.1.4. Modèle bivarié Trafic routier / trafic ferroviaire de fret

Dans l'écriture suivante³⁸, les séries, endogènes et exogènes, ont été filtrées préalablement par un filtre (I-B⁴).

$$\begin{aligned}
 sncf_t = & -0.13 sncf_{t-1} - 0.51 sncf_{t-2} - 0.43 sncf_{t-4} - 570.4 - 0.042 trm_{t-1} + 0.123 trm_{t-4} \\
 & - 31.9 ip1_t + 29.7 ip1_{t-4} + 72.1 ip3_t - 24.83 ip7_t + 10.23 ip7_{t-1} \\
 & + 16.05 ip931_t + 19.8 ip931_{t-2} - 23218.5 iprmtk_t + 11819.8 iprmtk_{t-1} \\
 & - 30977.4 iprmtk_{t-2} + 1668.6 iprout_{t-1} + 2516.7 iprout_{t-2} \\
 trm_t = & 0.23 trm_{t-1} + 0.09 trm_{t-2} - 0.38 trm_{t-4} + 139.2 + 0.63 sncf_{t-1} + 0.23 sncf_{t-2} + 0.68 sncf_{t-4} \\
 & + 32.23 ip1_t + 126.16 ip7_t - 33.9 ip7_{t-1} + 138.1 ip931_t - 52.4 ip931_{t-1} \\
 & - 21326.2 iprout_t + 11059.9 iprout_{t-1} - 7950.9 iprout_{t-2}
 \end{aligned}$$

route : écart-type résiduel=482.95 part de variance expliquée=85.3% (+11%)

fer : écart-type résiduel=178.26 part de variance expliquée=70.7% (+26.7%)

	1984		1992	
	<i>sncf</i>	<i>trm</i>	<i>sncf</i>	<i>trm</i>
<i>ip1</i>	-0.26	0.12	-0.35	0.10
<i>ip3</i>	0.53	0.22	0.66	0.18
<i>ip7</i>	-0.15	0.63	-0.23	0.60
<i>ip931</i>	0.19	0.73	0.26	0.62
<i>iprmtk</i>	-0.43	-	-0.39	-
<i>iprout</i>	0.206	-0.77	0.176	-0.42

Tableau 7.5- Elasticités apparentes pour le modèle bivarié trafic routier/ trafic ferroviaire de fret, sur 1980-1992

³⁸ Le modèle vectoriel peut aussi être écrit sous la forme matricielle simplifiée suivante :

$$Y_t = \phi'_1 Y_{t-1} + \dots + \phi'_p Y_{t-p} + \psi'_0 Z_t + \dots + \psi'_q Z_{t-q} + W$$

avec : Y le vecteur endogène à 2 composantes,

Z le vecteur exogène à 6 composantes,

$\Phi'(B) = I - \phi'(B) - \dots - \phi'_p(B)$ et $\Psi'(B) = I - \psi'(B) - \dots - \psi'_p(B)$ deux matrices de polynômes en B,

et $(I - B^s)$ le filtrage commun à toutes les séries, ou s est la saison des séries.

7.4.1.5. Analyse

Les quatre composantes représentatives de la production industrielle ressortent significatives des trafics intérieurs de fret³⁹. En faisant l'hypothèse d'effets additifs de ces quatre composantes, l'élasticité (apparente) à l'ensemble de la production industrielle serait supérieure à 1 pour la route, et inférieure à 1 pour le fer.

On notera des valeurs d'élasticités négatives à des composantes de la production industrielle, dans les modèles du fret ferroviaire, et ceci s'explique par la forte corrélation entre certaines variables explicatives utilisées simultanément dans le même modèle. Ainsi, l'élasticité du fret ferroviaire à la production des IAA ressort négativement en raison de la corrélation entre la recette unitaire et cette variable de production. L'élasticité du fret ferroviaire à la production du BTP ressort négativement en raison de la corrélation entre le prix routier et cette variable de production.

Le prix routier est significatif des deux trafics de fret⁴⁰, alors que la recette unitaire ferroviaire n'est significative que du fret ferroviaire. On notera que les ordres de grandeur des élasticités des deux trafics de fret à ces deux variables sont, en fin de période, du même ordre de grandeur, avoisinant -0.4, alors que l'élasticité additionnelle du fret ferroviaire au prix routier est inférieure, de l'ordre de 0.2 seulement.

7.4.1.6. Comparaison avec les modèles annuels

La comparaison avec les sorties des modèles sur données annuelles, en termes d'élasticité à un an, fait apparaître des phénomènes différents suivant la périodicité des données : on peut différencier les effets de court terme qu'on mesure sur données infra-annuelles, et les effets de tendance qu'on mesure sur données annuelles. C'est par exemple le cas de l'effet du prix routier sur le transport routier de marchandises, qui ressort avec une élasticité qui diminue de -0.85 à -0.5, ou de -0.5 à -0.3 sur la période, selon l'une ou l'autre approche. De plus, le produit moyen ferroviaire n'est significatif du fret ferroviaire qu'en rythme infra-annuel, en sus du prix routier - ce qui n'est pas le cas lorsqu'on le saisit en rythme annuel, où seul le prix routier est significatif.

³⁹ à l'exception de la production des industries agro-alimentaires *ip1* qui, à la limite de la significativité statistique pour le fret routier dans l'approche modale, a été regroupée avec la production d'ensemble dans la variable résiduelle *ip93*.

⁴⁰ En raison de sa corrélation avec la production du bâtiment, génie civil et agricole, la variable de prix routier ne ressort pas significative du fret ferroviaire lorsque cette variable de production l'est.

7.4.1.7. Performance

Au plan statistique, les modèles les plus performants sont obtenus dans l'approche bimodale, même si ce gain dont l'ordre de grandeur est faible comparativement aux modèles annuels doit être relativisé. Le léger gain obtenu en utilisant de tels modèles sur des années civiles n'est probablement pas sensible auprès des professionnels du secteur des transports, l'apport majeur restant pour ceux-ci la meilleure connaissance des mécanismes d'évolution à court terme du volume d'activité de leur secteur.

	Bivarié		Univarié	
	sncf	trm	sncf	Trm
1982	0.69%	1.58%	-0.02%	0.83%
1983	0.66%	1.23%	-0.49%	0.85%
1984	-0.11%	-1.16%	1.46%	-1.51%
1985	-0.83%	-0.34%	-0.89%	-0.58%
1986	0.64%	-0.83%	-0.61%	-1.75%
1987	0.01%	-0.19%	-0.58%	1.07%
1988	1.05%	-0.55%	2.13%	-0.03%
1989	-0.53%	1.02%	0.78%	0.30%
1990	1.70%	1.68%	2.23%	1.94%
1991	-3.03%	-1.00%	-2.35%	-1.07%
1992	-0.44%	-1.32%	-2.32%	-1.42%
moyenne	0.90%	0.93%	1.38%	1.05%

Tableau 7.6 - Performances estimatives à 1 an des modèles, sur 1982-1992

En revanche, les modèles sont moins interprétables dans l'approche bimodale que dans l'approche modale, si l'on cherche à disposer de mesures d'élasticité des trafics à leurs déterminants : les effets d'offre et de demande sont plus difficiles à séparer en raison du plus grand nombre de variables corrélées, et l'on ne peut que faire état d'élasticités apparentes. Ainsi, par exemple, l'effet de structure que subit la recette unitaire ferroviaire à la tonne-km se traduit pas une corrélation importante à certains secteurs de la production industrielle, ce qui interdit de l'utiliser comme une véritable variable exogène pour réaliser des scénarios prévisionnels. En l'absence de véritable indicateur de prix pour le fret ferroviaire, on peut préférer utiliser un modèle à seule variable de prix routier. Le même type de contrainte apparaît avec un autre exemple, lorsque l'on introduit des variables représentatives des échanges extérieurs dans les modèles des trafics globaux de fret, les importations étant corrélées aux variables de production.

7.4.2. Transports interurbains de voyageurs

Le deuxième exemple porte sur les transports interurbains de voyageurs, pour la route et pour le fer. Les modèles retenus sont utilisés pour analyser l'impact de fluctuations conjoncturelles importantes des variables d'offre et de demande qui ont été enregistrées en début d'année 1993, à la suite d'une refonte tarifaire proposée aux usagers de la SNCF, d'une modification du système de distribution ferroviaire et simultanément à une phase de récession économique.

7.4.2.1. Les variables

Les indicateurs modélisés sont : pour le mode routier, la circulation des véhicules sur routes nationales, sur autoroutes concédées, et sur l'ensemble du réseau routier national, et pour le mode ferroviaire, le trafic de voyageurs sur le réseau principal de la SNCF. Nous donnons ici les variables nécessaires au modèle présenté, à savoir :

<i>paer</i>	Circulation routière sur le réseau routier interurbain, mesurée en centaines de millions de véhicules-km, source SNRD
<i>vkpr</i>	Trafic ferroviaire de voyageurs sur le réseau principal de la SNCF, en centaines de voyageurs-km, source : recettes SNCF
<i>cfm</i>	Consommation finale des ménages, en milliards de francs 1980
<i>pibc</i>	Complément de la consommation finale des ménages dans le produit intérieur brut, , en milliards de francs 1980
<i>icarb</i>	Prix des carburants utilisés par les voitures, en francs constants 1980, source INRETS
<i>ipeagetl</i>	Recette unitaire de péage autoroutier, tous véhicules confondus, en centimes constants 1980
<i>ipmfer</i>	Recette unitaire du trafic de voyageurs sur le réseau principal de la SNCF, en francs constants 1980 au voyageur-km, source SNCF
<i>lgac</i>	Longueur du réseau autoroutier concédé, en km, source SNRD
<i>trkmTGVs</i>	Offre en trains-kilomètres sur le réseau TGV Sud-Est de la SNCF, en milliards de trains-km, source SNCF

Tableau 7.7 - Variables utilisées dans l'exemple des modèles de transports interurbains de voyageurs

7.4.2.2. Modèle univarié de la circulation routière

Meilleur modèle avec la *cfm* :

$$(I - B^4)(I - 0.24 B^1)_{(0.16)}[paer_t - Z_t] = (I - 0.39 B^4)_{(0.16)}\varepsilon_t + 1.18$$

$$avec les corrections Z_t = 0.28_{(0.15)} cfm_t + 0.54_{(0.13)} cfm_{t-1} - 36.6_{(18.9)} icarb_t - 4.03_{(1.8)} peagetl_t$$

écart-type résiduel=5.67

part de variance expliquée=46.3%

Meilleur modèle avec le pib :

$$(I - B^4)(I - 0.10 B) [paer_t - Z_t] = (I - 0.64 B^4) \varepsilon_t + 1.01$$

$$aveclescorrections Z_t = 0.36 pib_t + 0.13 pib_{t-1} - 65.6 icarb_t - 3.3 peagetl_t$$

écart-type résiduel=5.23

part de variance expliquée=54.3%

7.4.2.3. Modèle univarié du trafic SNCF réseau principal

$$(I - B^4)(I + 0.13 B)(vkrp_t - Z_t) = (I - 0.64 B^4) \varepsilon_t + 52.6$$

$$aveclescorrections Z_t = 9.44 cfm_t + 2556.74 icarb_t - 44768.8 ipmfer_t + 11094.5 ipmfer_{t-1} + 163.8 trkmtgvs_t$$

écart-type résiduel=290.9

part de variance expliquée=43.9%

7.4.2.4. Modèle bivarié Circulation routière/ Trafic SNCF réseau principal

Dans l'écriture suivante, les séries, endogènes et exogènes, ont été filtrées préalablement par un filtre $(I-B^4)$.

$$paer_t = -0.37 paer_{t-1} - 0.17 paer_{t-2} - 0.34 paer_{t-4} - 0.0046 vkrp_{t-1} - 0.0032 vkrp_{t-2} - 0.0022 vkrp_{t-4}$$

$$+ 0.84 cfm_t + 0.34 cfm_{t-1} + 0.47 cfm_{t-4} - 40.95 icarb_t + 45.4 icarb_{t-1} - 19.4 icarb_{t-2} - 18.7 icarb_{t-4}$$

$$+ 0.43 pibc_t - 3.23 trkmtgvs_t + 1.91$$

$$vkrp_t = -0.06 vkrp_{t-1} + 0.20 vkrp_{t-2} - 0.48 vkrp_{t-4} + 2.9 paer_{t-1} + 8 paer_{t-2} - 16.2 paer_{t-4} + 24.1 cfm_t$$

$$- 19.5 cfm_{t-1} + 21.4 cfm_{t-2} + 2888.3 icarb_t + 278.5 icarb_{t-1} + 2086.6 icarb_{t-4} - 45513.8 ipmfer_t$$

$$+ 26233 ipmfer_{t-1} - 24382.7 ipmfer_{t-4} - 0.59 lg ac_t + 116.5 trkmtgvs_t + 240.6$$

route : écart-type résiduel=3.81 part de variance expliquée=76.8%

fer : écart-type résiduel=247.4 part de variance expliquée=59.4%

7.4.2.5. Un exemple de mise en œuvre

En utilisant différents modèles existant, calés sur la période 1980-1992, et en simulant l'évolution des trafics sur la période 1993-1995, à variables exogènes connues, on tente d'expliquer le " choc " qu'a connu le trafic ferroviaire de voyageurs en début d'année 1993, au moment de la refonte du système tarifaire et de la mise en place du nouveau système de distribution Socrate.

La différence entre les réalisations et les prévisions de trafic, doit permettre d'apprécier l'effet de ce choc. De fait, ces prévisions de trafic ont été réalisées a posteriori, donc à variables

d'offre connues, qui ne pouvaient pas intégrer les nouvelles caractéristiques de l'offre. En effet, l'indicateur de recette unitaire reproduit mal l'évolution tarifaire réelle des différents produits qui ont connu une augmentation supérieure à celle de la taxe kilométrique de base, et notamment des réservations sur les trains TGV dont l'augmentation a été plus marquée; par ailleurs les modèles ne comportent pas de variable représentative de la qualité de l'offre ferroviaire qui a souffert de la mise en service du nouveau système de distribution.

Pour le trafic de voyageurs sur le réseau principal de la SNCF, le choc du début 1993 apparaît enrayé en 1995.

La perte de trafic non expliquée par les modèles serait de l'ordre de 6% en 1993 (-7,9% en réel comparativement à -2,3% par modélisation), de l'ordre de 4% supplémentaires en 1994 (+1,4% comparativement à 5,4%), alors qu'en 1995 les deux évolutions sont très voisines (+1,7% ,⁴¹ comparativement à +1,2%).

Le même exercice de comparaison des prévisions à trois ans avec les réalisations, pour la circulation routière interurbaine, a fait apparaître, un report modal important sur le réseau routier en 1993, atténué en 1994 et nul en 1995.

Pour l'année 1993, l'effet du report modal dû à la désaffectation des usagers de la SNCF serait de l'ordre de 3,5%⁴² sur l'ensemble du réseau routier (+3,3% comparativement à -0,2%), de l'ordre de 3% sur les seules autoroutes concédées (+4,7% comparativement à +1,6%), de l'ordre de 4% sur les autoroutes concédées pour les seuls voyageurs (+6,6% comparativement à +2,3%) et de l'ordre de 1% sur routes nationales (+1,4% comparativement à +0,3%).

En 1994, le report modal, qui reste le seul facteur explicatif des écarts entre évolutions réelles et évolutions simulées, serait faible (+3,5% comparativement à +3% sur l'ensemble du réseau routier, et +3,8% comparativement à +3,1% sur le réseau autoroutier concédé, +3,9% comparativement à +3,5% pour les seuls voyageurs sur le réseau autoroutier concédé, enfin +2,4% comparativement à +2,6% sur routes nationales.

Pour 1995, les évolutions réelles et estimées sont voisines sur l'ensemble du réseau.

⁴¹ si l'on corrige les réalisations de l'effet des grèves de novembre et décembre 1995 dans les transports publics

⁴² les évolutions enregistrées intègrent aussi un supplément mécanique de croissance du fait de l'impact de la grève des routiers de juillet 1992 sur le niveau des trafics de l'année 1992

7.5. Actualisation

La modélisation multivariée des trafics terrestres de voyageurs et de marchandises produite en 1997 a été actualisée en 2000 et 2002, en rythme trimestriel et semestriel et pour un petit nombre d'indicateurs de trafic (*Rapports de convention SES-Inrets n° 00/5814-15*) en vue de sa mise en œuvre au SES.

Ce travail s'est fait en deux étapes :

1/ Actualisation de la base de données trimestrielle à partir de la base mensuelle, sur 1980-1998. Actualisation du modèle et *choix d'une autre spécification économétrique (Log-Log* qui est la meilleure pour la prévision à un pas⁴³). Mise en œuvre sur 1998 et 1999 : analyse de l'activité 1998 et anticipation de l'activité 1999, sur la base d'un jeu d'hypothèses retenu à la fin 1999.

2/ Actualisation de la base de données trimestrielles sur 1980-1999, en intégrant les agrégats macro-économiques des comptes nationaux trimestriels en (nouvelle) base 1995, et passage en base semestrielle. Actualisation du modèle trimestriel et calibrage en format semestriel. Transfert sous E-views des modèles estimés sous SAS, en trimestriel et en semestriel.

Après actualisation de la base de données sur les vingt ans, et du fait des changements induits par la comptabilité nationale trimestrielle en base 1995, les modèles ressortent modifiés (non significativité de certaines variables et partage différent des effets, cf. tableaux 7.10 et 7.11).

Les modèles semestriels ont la même logique que les modèles trimestriels, mais avec une meilleure performance qui tient à la périodicité semestrielle.

Enfin la relance des modèles sous E-views n'induit pas de modification des résultats obtenus sous SAS.

7.6. Conclusion

Le premier objectif qui consistait à prendre en compte sur une base infra-annuelle les principaux déterminants représentatifs de l'offre et de la demande de transport dans un modèle explicatif des trafics agrégés de voyageurs et de marchandises, a globalement bien été

⁴³ La spécification *Log-Log* a été retenue lors de l'actualisation, de préférence à deux autres spécifications (*Log-lin* et *lin-lin*,) comme la meilleure sur la base d'un critère de performance à un pas. Les paramètres des variables exogènes représentent les élasticités apparentes de l'endogène aux exogènes, supposée constante sur la période. Des valeurs d'élasticités à court terme du trafic ferroviaire de voyageurs à la consommation et aux prix, estimées par une méthode de cointégration, sont données dans (Bergel et al, 1995)

atteint. De plus, avec le critère de performance retenu, les estimations agrégées sur une année civile sont légèrement meilleures que les estimations obtenues avec les modèles annuels.

Le second objectif qui consistait à tirer des enseignements de la comparaison des résultats obtenus avec ceux des modèles annuels n'a été que partiellement atteint. En effet, les corrélations entre certaines variables explicatives rendent difficile l'interprétation des valeurs d'élasticités pour ce groupe de variables. Ces corrélations sont plus nombreuses dans l'approche bimodale que dans l'approche modale, et on privilégiera en fonction des objectifs un modèle bivarié pour son meilleur ajustement, ou un modèle univarié pour sa plus grande interprétabilité.

L'actualisation des modèles précédents a conduit à retenir une spécification différente, à élasticités constantes. Ces modèles sont apparus robustes lors du passage d'un format trimestriel à un format trimestriel, et à un changement de logiciel (passage de SAS à E-views). En revanche, l'utilisation des nouveaux agrégats de la comptabilité nationale a induit des modifications notables dans les modèles antérieurs.

7.7. Discussion et prolongements

Le point de débat central est la limitation de modèles agrégés à reproduire une réalité complexe de manière trop simplifiée et à fournir des paramètres interprétables. Les causes en sont diverses, et liées, puisqu'il s'agit tout-à-la fois d'utiliser les méthodes de modélisation appropriées à partir de représentations économiques utilisant les facteurs de demande adéquats, et de disposer de bases de données appropriées pour mesurer les indicateurs et les facteurs de demande.

Il ressort clairement qu'une approche agrégée est limitée dans son objectif, puisque les spécificités des marchés, tant pour les transports de personnes que pour les trafics de fret, ne peuvent être pris en compte de manière satisfaisante que dans une optique relativement désagrégée. La formulation économique doit s'appuyer sur une segmentation des déplacements de voyageurs et sur une approche sectorielle des transports de marchandises, sous réserve bien entendu que les connaissances sur les bases de données des facteurs de demande progressent. Il s'agit par exemple d'améliorer la mesure des variables de qualité de l'offre (vitesse des trains en rythme infra-annuel, décomposée selon le type de réseau ferroviaire), et celle des variables de prix (ou de recettes unitaires désagrégées) tant pour les transports de voyageurs que pour les trafics de fret. De plus, la validité des données utilisées

pour mesurer les facteurs de demande doit être précisée - la mesure des agrégats de comptabilité nationale pouvant à elle seule, comme nous l'avons constaté en intégrant les séries trimestrielles de la nouvelle base 1995, occasionner des changements sensibles des paramètres estimés dans les modèles.

L'approche économétrique multivariée, c'est-à-dire multi-modale, ressort inappropriée à l'interprétation des paramètres estimés, en raison de la corrélation entre variables exogènes fortement corrélées. L'approche économétrique modale, plus adaptée à l'interprétation des paramètres, est intéressante dans la mesure où elle fait ici ressortir une dynamique de très court terme qui ne peut pas être saisie sur une base annuelle. C'est par exemple le cas de la variable de prix (recette unitaire) ferroviaire qui ressort comme significative du trafic domestique de fret ferroviaire - les contrats signés entre les chargeurs et la SNCF sont bloqués pour un an, terme à partir duquel les prix pratiqués peuvent se rapprocher de ceux du mode routier - même en étant mesurée de manière très agrégée. Pour autant, la prise en compte de deux variables mesurant nos échanges extérieurs dans des modèles explicatifs des trafics de fret globaux (champ intérieur et international) ramène à la difficulté d'interpréter des paramètres relatifs à des variables corrélées, notamment d'importations et de production, même dans une approche modale.

Enfin, les modèles à élasticité constante, qui fournissent une bonne estimation à un pas (ce qui était ici un objectif majeur pour une estimation à court terme), ne sont pas appropriés pour représenter les transformations que l'on observe sur longue période. La séparation des mouvements de court terme et de long terme est d'ailleurs l'orientation prise en modélisation économétrique pour appréhender les évolutions à long terme. De nombreux travaux visant à appliquer les techniques de cointégration au secteur du transport, et à l'analyse de l'évolution des trafics de voyageurs et de marchandises pour différents modes de transport ont été menés ces dernières années (Meyer, 1998, Muzereau, 2004, Lenormand, 2007). C'est dans cette direction méthodologique que les efforts doivent porter.

Tableau 7.8 : Les données de trafic en 1992

Voyageurs								
	Routes nationales	Autoroutes concédées	Autoroutes non concédées	Réseau routier national	Réseau principal SNCF	Lignes rapides	Lignes régionales	
Trafic annuel	78,7	45,44	26,88	151,01	52,93	47,21	5,72	
moy. mensuelle	7,89	3,79	2,24	12,58	4,41	3,93	0,48	
Marchandises								
	Domestique		International	Fret routier total	Fret ferroviaire total	Domestique	International	
Trafic annuel	101,68		18,03	119,71	50,37	29,49	20,89	
moy. mensuelle	8,47		1,5	9,98	4,2	2,46	1,74	

Unités : *Milliards de véhicules-kms et milliards de voyageurs-kms pour le transport de voyageurs, Milliards de tonnes-kms pour le trafic de marchandises.*

Tableau 7.9 : La structure des modèles de demande

Modes	Routier	Ferroviaire
Variables modélisées	Trafic (voyageurs, marchandises)	Trafic (voyageurs, marchandises)
Variables explicatives		
Croissance économique	Consommation des ménages Produit intérieur brut Production industrielle	Consommation des ménages Produit intérieur brut Production industrielle
Réseau	Longueur du réseau autoroutier	Trains-kilomètres ou Vitesse des trains
Prix	Prix des carburants Péage autoroutier Prix du transport routier	Recette unitaire (voyageurs, fret wagons)

Tableau 7.10: Modèles de transports interurbains de voyageurs, bases trimestrielle et semestrielle, sur 1980-1999 (Sorties SAS et E-views)

Endogène	R ²	Elasticités				
		PIB	ICARB	PEAGETL	DCM	IPMFER
Modèles trimestriels sous SAS (maximum de vraisemblance)						
PAAC	45,30%	1,17	-0,24	-0,53		
PAER	41,30%	0,79	-0,13	-0,13		
VKRP (*)	50,70%		0,16		1,05	-0,47
Modèles semestriels sous SAS (maximum de vraisemblance)						
PAAC	69,40%	1,02	-0,24	-0,47		
PAER	61,20%	0,46	-0,08	-0,23		
VKRP (*)	59,50%		0,18		1,3	-0,78
Modèles trimestriels sous E-views (moindres carrés)						
PAAC	46,20%	1,20	-0,24	-0,52		
PAER	40,70%	0,77	-0,13	-0,14		
VKRP (*)	55,80%		0,16		0,56	-0,18
Modèles semestriels sous E-views (moindres carrés)						
PAAC	69,70%	1,05	-0,25	-0,45		
PAER	61,10%	0,45	-0,08	-0,26		
VKRP (*)	59,90%		0,19		1,18	-0,83

Seuil de significativité pour tous les paramètres: t-ratio supérieur à 1 en valeur absolue

(*)La série du trafic ferroviaire de voyageurs est corrigée de la grève de novembre-décembre 1995

Avec : PAAC et PAER les parcours effectués par l'ensemble des véhicules sur les autoroutes concédées et sur l'ensemble du réseau routier national

VKRP le trafic de voyageurs sur le réseau principal de la SNCF,

PIB le produit intérieur brut

DCM la dépense de consommation des ménages,

ICARB le prix des carburants voiture,

IPEAGETL le tarif de péage autoroutier pour l'ensemble des véhicules

IPMFER la recette unitaire sur le réseau principal de la SNCF

(cf. la description précise des variables donnée dans le Tableau 7.7).

Tableau 7.11: Modèles de trafics de marchandises, bases trimestrielle et semestrielle, sur 1980-1999 (Sorties SAS et E-views)

Endogène	R ²	Elasticités				
		IP3	IP7	IP9	IPRMTK	IPROUT
Modèles trimestriels sous SAS (maximum de vraisemblance)						
TRMTOTR	74,5%			1,44		-0,50
TRMR	73,5%			1,46		-0,49
SNCF TOT (**)	82,6%	1,56			-0,50	0,43
SNCF (**)	66,2%	0,92			-0,20	
Modèles semestriels sous SAS (maximum de vraisemblance)						
TRMTOTR	73,2%		0,44	0,6		-0,61
TRMR	71,5%		0,41	0,56		-0,52
SNCF TOT (**)	90,5%	1,65			-0,47	0,42
SNCF (**)	73,4%	0,9			-0,21	
Modèles trimestriels sous E-views (moindres carrés)						
TRMTOTR	77,2%			1,17		-0,41
TRMR	75,4%			1,20		-0,30
SNCF TOT (**)	84,0%	1,53			-0,48	0,44
SNCF (**)	68,0%	0,93			-0,20	
Modèles semestriels sous E-views (moindres carrés)						
TRMTOTR	78,3%		0,39	0,45		-0,37
TRMR	77,0%		0,37	0,42		-0,27
SNCF TOT (**)	95,9%	1,62			-0,52	0,40
SNCF (**)	75,2%	0,88			-0,21	

Seuil de significativité pour tous les paramètres: t-ratio supérieur à 1 en valeur absolue

(**) Les séries de fret ferroviaire sont corrigées de la grève de novembre-décembre 1995

Avec : TRMTOTR et TRMR le transport routier de marchandises, total et national,
 SNCTTOT et SNCF le fret ferroviaire wagons, total et national,
 IP7 la production du BTP,
 IP9 la production industrielle hors BTP et IP3 la production des biens intermédiaires,
 IPRMTK la recette unitaire wagons et IPROUT le prix routier en intérieur et en zone longue
 (cf. la description précise des variables donnée dans le Tableau 7.1).

Partie III : Applications au risque routier

Chapitre 8.

Les modèles agrégés de risque routier

Nous nous intéressons ici aux modèles agrégés de risque routier, à périodicité annuelle, mensuelle ou journalière. Les variables modélisées représentant les indicateurs de risque sont dans tous les cas agrégées au niveau d'un territoire, d'un réseau, ou suivant une typologie d'accidents corporels ou de victimes.

Nous exposons d'abord le cadre général de la modélisation agrégée du risque routier : historique de la modélisation, cadre formel, évolution des formulations économétriques. Nous donnons ensuite la problématique retenue dans le cadre de la thèse, qui vise à évaluer les effets de nature transitoire, de nature climatique et calendaire, qui se manifestent sur le risque routier ; enfin nous indiquons les types de travaux menés avec des visées différentes sur les mêmes indicateurs agrégés d'accidents corporels et de victimes, qui vont être détaillés dans les chapitres 9 à 12.

8.1. Historique de la modélisation de l'insécurité routière

Le premier modèle de l'insécurité routière est reconnu être celui de Smeed (1949) ; il relie le nombre annuel de décès dans des accidents de la circulation routière au nombre de véhicules motorisés et à la population correspondante. Bien que la loi de Smeed ne s'ajuste plus aux données depuis les années 1970, l'on voit déjà apparaître ici de manière sous-jacente la variable de mesure de l'exposition au risque qu'est le volume de trafic - plus précisément le nombre de véhicule-kilomètres parcourus par une population sur un réseau routier.

Les premières mesures visant à réduire l'insécurité routière ont été prises en 1968 aux Etats-Unis, et Peltzmann (1975) suscite un débat important en mettant en cause leur efficacité. Il développe une théorie du comportement du conducteur, qui réaliserait un arbitrage entre un type de conduite, caractérisé par un indice de prise de risque, et une probabilité de décès : cette *adaptation au risque*⁴⁴ de la part des conducteurs expliquerait l'inefficacité des mesures de sécurité routière.

⁴⁴ La théorie de l'homéostasie du risque (ou adaptation à un niveau de risque) a été formalisée par la suite par Wilde (1982)

Suite à ce débat, les gains constatés au début des années 1980 aux Etats-Unis et en Grande Bretagne ont amené à s'interroger sur leurs causes : pouvait-on attribuer ces gains à l'effet des politiques de sécurité routière, ou à des facteurs exogènes à la sécurité routière ?

Une *approche systémique* s'est imposée : elle consiste à appréhender le système sécurité routière dans sa globalité, en prenant en compte l'ensemble des déterminants du risque routier. Des modèles globaux sur données agrégées ont été développés, avec un double objectif : *évaluer l'effet des politiques de sécurité routière, et estimer simultanément l'impact des autres déterminants.*

Oppe et Koornstra (1990), en modélisant le nombre de tués comme le produit du taux de tués⁴⁵ et du volume de trafic, parviennent à expliquer ce renversement de tendance du nombre de décès sur la route observé dans les années 1970 dans les pays développés.

La tendance du taux de tués est modélisée comme une fonction exponentielle décroissante du temps. Ce modèle met en évidence *l'importance du trafic en tant que facteur d'exposition au risque, et son effet multiplicatif sur le nombre de tués.* Le modèle proposé pour le taux de tués a ensuite été suivi d'une formulation plus générale, dans laquelle le nombre de tués est modélisé directement comme une fonction du trafic, et n'est plus contraint à être proportionnel au trafic⁴⁶

L'approche systémique a permis l'émergence d'*un ensemble de variables associées à des facteurs de risque*, grâce à un grand nombre de modèles sur données agrégées, développés sur coupes transversales et sur séries temporelles à périodicité annuelle et mensuelle (Hakim et al. 1991). Outre le volume de trafic comme mesure de l'exposition au risque, ou encore le parc de véhicules, nous citerons l'activité économique mesurée par exemple par le revenu des ménages, le prix des carburants, la structure de la population mesurée par exemple par la part des jeunes conducteurs, et des mesures réglementaires telles que les limitations de vitesse, l'instauration du contrôle technique ou de l'âge d'accès à la boisson alcoolisée.

Wirkowsky et Heckard (1977) ont proposé une estimation des effets du choc pétrolier et de la limitation de vitesse de 1974 sur le taux de tués au Texas, et Wagenaar et Maybe (1986) ont cherché à estimer la relation entre âge d'accès à la boisson alcoolisée et l'insécurité, toujours

⁴⁵ nombre de décès par véhicule-kilomètre

au Texas. Les deux exemples suivants montrent que des démarches de modélisation différentes peuvent produire les mêmes évaluations. En Grande Bretagne, Scott (1986) estime l'effet sur l'accidentologie du choc pétrolier de 1974 et de la limitation de vitesse en rase campagne, tout en prenant en compte le trafic, le prix des carburants, l'influence du climat par le biais de la température et de la hauteur de pluie, et le nombre de jours ouvrables. Toujours en Grande Bretagne, Harvey et Durbin (1986) proposent une estimation de l'impact de la loi sur le port de la ceinture de sécurité, tout en intégrant l'effet du trafic et du prix des carburants ; *le caractère novateur de leur approche structurelle*, fondée sur l'estimation de la tendance et de la saisonnalité en tant que composantes aléatoires a suscité un débat important. Scott, en utilisant des modèles autorégressifs sur les mêmes données, a confirmé les estimations proposées par Harvey et Durbin pour l'impact de la loi sur le port de la ceinture de sécurité.

La modélisation structurelle proposée par Harvey, que nous développons plus avant en 8.3, a été utilisée par la suite de manière comparative sur les données agrégées de plusieurs pays européens. L'impact des principales mesures de sécurité routière sur les niveaux et les pentes des tendances des nombres annuels de tués a été identifié, avec le volume de trafic comme seul déterminant du nombre de tués (Lassarre, 1996 et 2001). Plus récemment, la modélisation structurelle a été développée dans un cadre multivarié : modéliser simultanément des données d'accidents et de mobilité permet d'estimer le risque latent correspondant à ces données (Bijleveld et al., 2008).

Comme alternative à la modélisation structurelle, *la démarche la plus répandue consiste à utiliser des modèles autorégressifs* ou des modèles s'y ramenant. On modélise les nombres mensuels d'accidents ou de tués en éliminant la tendance et la saisonnalité, mais on intègre un petit nombre de variables mesurant : le trafic, l'activité économique et le prix des carburants, les conditions climatiques et la configuration mensuelle du calendrier, ainsi que des variables d'intervention qui permettent de d'évaluer les effets de mesures de sécurité routière et d'événements exceptionnels (COST 329, 2004).

⁴⁶ De manière similaire, la modélisation directe du taux d'accidents (nombre d'accidents corporels par véhicule-kilomètre) est relayée par une modélisation du nombre d'accidents par une fonction non linéaire du trafic. Cette formulation trouve d'ailleurs sa principale justification dans les résultats des études microscopiques d'accidents.

La forme la plus avancée de modèle autorégressif explicatif du risque routier est due à Gaudry (1984) : celui-ci propose *une représentation structurée du système de sécurité routière* avec un modèle sur données mensuelles agrégées, qui est *explicatif, à trois niveaux et à forme fonctionnelle plus générale* que celle des modèles précédents.

Gaudry formule le nombre de victimes (tués, blessés graves et blessés légers) comme le produit du kilométrage, du taux d'accident (nombre d'accidents par véhicule-kilomètre) et du taux de gravité (nombre de tués, blessés graves et blessés légers par accident). En utilisant un très grand nombre de variables explicatives, qui sont les mêmes à chacun des 3 niveaux, il parvient à prendre en compte les effets directs et indirects - via le kilométrage - des déterminants du risque d'accident et de sa gravité. Enfin, en utilisant une transformation de Box-Cox sur les variables, il autorise des effets non strictement multiplicatifs.

Cette démarche de modélisation, baptisée DRAG (pour demande routière, accident, gravité) a été reprise par d'autres chercheurs, en Europe et aux Etats-Unis, qui l'ont mise en œuvre tout en l'adaptant aux spécificités de leur pays (Gaudry, Lassarre (Eds.), 2000). Sur données françaises, par exemple, un niveau supplémentaire représentant le comportement à risque mesuré par la vitesse pratiquée sur le réseau interurbain a pu être constitué (Jaeger, Lassarre, 1999).

Pour être complète, mentionnons enfin la modélisation des données journalières d'insécurité routière, fort peu répandue. De fait, la grande variabilité des statistiques d'accidents et de victimes rend très délicate l'analyse des évolutions de court terme. Les modèles sur données journalières répondent à un double objectif : d'une part *la connaissance des effets transitoires*, dus à une météorologie atypique ou à une configuration exceptionnelle du calendrier, et d'autre part *l'estimation d'une tendance journalière corrigée de ces effets locaux*, sur laquelle on puisse détecter des inflexions. Citons le modèle de désaisonnalisation de données journalières d'insécurité routière développé sur données françaises (Bergel et al, 1995) qui est discuté dans le chapitre 9. La tendance mensuelle reste cependant la référence pour le suivi conjoncturel : très répandue dans de nombreux autres secteurs, l'analyse des statistiques mensuelles désaisonnalisées commence maintenant à être exploitée par les décideurs publics de la sécurité routière (La sécurité routière en France : bilan annuel de l'année, ONISR, 2006).

8.2. Le cadre

Nous rappelons ici le cadre retenu pour quantifier l'influence sur le risque routier des différents facteurs relatifs au système de transport, à la mobilité, à l'économie ou à la gestion de la sécurité routière (Lassarre, 1994). Nous nous intéressons aux modèles de séries temporelles agrégées, à périodicité annuelle, mensuelle ou journalière. Les variables modélisées sont dans tous les cas agrégées au niveau d'un territoire, d'un réseau, ou suivant une typologie d'accidents ou de victimes.

8.2.1. Vers des modèles explicatifs

Comme nous l'avons vu, le suivi temporel ne se limite plus depuis le début des années 1980 à une approche descriptive traditionnelle, de simple constat, mais vise à quantifier l'influence des facteurs explicatifs de la fréquence de la gravité des accidents tout en évaluant l'efficacité des mesures de sécurité routière. Ainsi, à des modèles *descriptifs* ont succédé des modèles *explicatifs*, ou à *variables explicatives*, construits sur la base d'une formulation économique riche, avec une spécification économétrique plus élaborée.

8.2.2. Le schéma de production du risque

L'analyse du risque se fonde sur la triade *exposition/accident/victime*. Dans ce schéma, on définit les indicateurs et leurs déterminants à chacun de ces 3 niveaux.

8.2.2.1. Les indicateurs

La mesure usuelle de l'exposition au risque est un indicateur de mesure du volume de trafic : on retient le kilométrage, ou parcours effectué, mesuré en *nombre de véhicule-kilomètres* parcourus sur un réseau routier.

On utilise habituellement *le taux d'accident* (nombre d'accidents corporels pour un milliard de véhicule-kilomètres) pour mesurer le risque d'accident sur un réseau ; pour contourner l'hypothèse de proportionnalité du nombre d'accidents au trafic, on mesure également le risque d'accident par *un nombre absolu d'accidents*, que l'on fait ensuite dépendre d'une fonction non linéaire du kilométrage⁴⁷.

⁴⁷ La même remarque vaut pour le risque d'être tué (ou encore taux de tués, soit le nombre de tués pour un milliard de véhicule-kilomètres).

Enfin, les indicateurs de mesure de la gravité de l'accident sont *les taux de gravité*, c'est-à-dire les nombres de victimes (tués, blessés graves, blessés légers) par accident ; on peut aussi préférer mesurer directement *le nombre de victimes*, mais en faisant là aussi dépendre le nombre de victimes du nombre d'accidents corporels, ou directement du trafic.

Notons à ce stade que les nombres absolus d'accidents et de victimes sont aussi considérés comme des indicateurs de risque et de gravité.

8.2.2.2. Les déterminants

On classe les déterminants en *facteurs internes au système de transport* d'un côté, - relatifs au véhicule, au conducteur et à l'infrastructure ; et en *facteurs externes* de l'autre, représentant l'environnement, et relatifs aux systèmes climatique, économique, démographique, et étatique (Gaudry, Lassarre (Eds.), 2000).

8.3. Applications majeures

Une revue bibliographique des modèles de séries chronologiques proposés pour l'analyse de l'évolution du risque routier nous amène à distinguer deux catégories de modèles. Les modèles descriptifs et les modèles à visée explicative. Comme déjà indiqué en 1.1, dans le cadre de la thèse les modèles descriptifs sont définis comme n'ayant aucune variable explicative autre que le temps, et les modèles explicatifs (ou à variables explicatives) comme des modèles descriptifs auxquels des variables exogènes ont été ajoutées. Ceux-ci constituent donc une extension des premiers.

Dans cette section, nous considérons successivement ces deux types de modèles, tels qu'ils ont été utilisés pour l'analyse agrégée de l'évolution de l'insécurité routière en Europe depuis le début des années 1980. La structure de base des modèles à laquelle nous ferons référence dans cette section a déjà été exposée en 1.1.1.

8.3.1. Modèles descriptifs : Structure déterministe ou stochastique ?

Les modèles purement descriptifs ont principalement été utilisés pour modéliser un indicateur de sécurité routière : le taux de tués. Il s'agit de modèles de décomposition dont l'objectif consistait à ajuster la tendance en fonction du temps. La décomposition tendance/résidu retenue sur une base annuelle a été étendue à une décomposition tendance/saisonnalité/résidu sur une base mensuelle. La composante tendancielle, ainsi que la composante saisonnière, est déterministe ou stochastique.

Ainsi, sur données annuelles, un exemple de modèle déterministe est fourni par Oppe (1991), qui retient une tendance exponentielle décroissante pour le taux de tués R_t (nombre de tués par milliard de véhicule-kilomètres) :

$$R_t = \exp(\alpha t + \beta)$$

avec : $R_t = \frac{F_t}{V_t}$,

F_t le nombre de tués,

et V_t le volume de trafic.

La forme proposée pour la tendance du taux de tués R_t peut être élargie, et une transformation de type puissance sur le trafic peut être retenue, pour rendre compte de la non proportionnalité du nombre de tués au trafic, le paramètre supplémentaire η représentant alors l'élasticité du nombre de tués au trafic:

$$\frac{F_t}{V_t^\eta} = \exp(\mu_t)$$

Dans les deux cas précédents, la tendance du taux de tués est modélisée de manière déterministe, en fonction du temps, la fonction la plus simple étant la fonction linéaire $\mu_t = \alpha t + \beta$, comme dans le premier cas. Mais la tendance peut aussi être aléatoire, auquel cas un terme d'erreur spécifique vient s'ajouter au modèle pour prendre en compte ce caractère aléatoire. Plus précisément, il y aura autant de termes d'erreur additionnels dans le modèle que de composantes, ou de paramètres, aléatoires.

Une forme stochastique a ainsi été proposée par Lassarre (1996) pour la fonction temporelle μ_{tt} qui devient linéaire localement, soit en complétant la formulation du modèle structurel de base proposé par Harvey:

$$\text{Log}F_t = \eta \text{Log}V_t + \mu_t + \varepsilon_t$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$

$$\beta_t = \beta_{t-1} + \xi_t$$

avec : β la pente de la tendance μ ,

ε , η , ζ des bruits blancs de variances σ_ε^2 , σ_η^2 et σ_ζ^2 , non corrélés entre eux.

Sur données mensuelles, la formulation s'enrichit d'une composante saisonnière, qui peut aussi être déterministe ou stochastique. En fait, le nombre plus important de données disponibles sur une base mensuelle autorise l'estimation de paramètres supplémentaires, c'est-à-dire l'utilisation de variables exogènes supplémentaires - nous l'exposerons dans la partie suivante.

Comme nous venons de le voir, un modèle descriptif du taux de tués peut être considéré comme un modèle explicatif du nombre absolu de tués, ayant pour seule variable explicative le volume de trafic. Cette forme de modèle explicatif à une seule variable exogène s'est ensuite enrichie par l'introduction de variables supplémentaires, en nombre plus ou moins important. On tient en fait pour modèles explicatifs les modèles prenant en compte un nombre plus important de facteurs de risque. Nous en donnons quelques exemples ci-après.

Remarquons au passage que la même formulation que celle qui vient d'être exposée pour modéliser le nombre de tués peut-être retenue pour un modèle explicatif du nombre d'accidents, en fonction du trafic puis de variables supplémentaires.

8.3.2. Modèles explicatifs : quelles variables pour mesurer les facteurs de risque ?

Citons d'emblée l'ensemble des variables émergeant de la revue bibliographique.

D'abord les variables qui mesurent l'exposition au risque (sous la forme de la population, du parc de véhicules, du trafic ou des consommations de carburants lorsque celui-ci n'est pas mesuré), puis les variables économiques représentatives de l'offre et de la demande de transport (activité économique, prix des carburants, consistance du réseau), puis les variables climatiques et les variables calendaires – des indicatrices témoins de jours particuliers de l'année. Des variables d'intervention sont aussi utilisées pour modéliser l'effet d'une mesure réglementaire de sécurité routière, ou d'un événement spécifique lié à la sécurité routière - cette partie du modèle pouvant à elle seule constituer l'objectif de la modélisation. Il reste à préciser, pour chaque groupe de variables, la manière dont elles sont introduites dans la spécification économétrique.

Comme nous l'avons vu pour la modélisation du taux de tués, l'exposition au risque mesurée par le nombre de véhicule-kilomètres occupe la première place parmi l'ensemble des variables que l'on retient pour modéliser les nombres de tués, et de manière générale les nombres d'accidents et de victimes. La spécification est de type Log-Log, ce qui traduit un effet multiplicatif du trafic, avec une élasticité du nombre d'accidents et de victimes au trafic qui est supposée constante au cours du temps. On formule la même hypothèse, d'un effet multiplicatif, lorsqu'on utilise d'autres variables de mesure de l'exposition au risque – la consommation de carburants, le parc de véhicules ou la population –, de l'activité économique – le revenu, la production industrielle, le taux de chômage –, ou de l'offre de transport – la consistance du réseau, le prix des carburants, le péage autoroutier.

Les variables climatiques interviennent de manière semi multiplicative, la spécification étant alors Log-linéaire, et il en est de même pour les variables indicatrices des jours de l'année représentatifs d'une configuration calendaire particulière.

Les autres variables d'intervention qui sont utilisées pour modéliser l'effet d'une mesure réglementaire de sécurité routière, ou d'un événement spécifique lié à la sécurité routière, interviennent de la même manière.

Nous présentons maintenant successivement des modèles de décomposition avec variables explicatives, à structure déterministe ou stochastique, puis les modèles ARIMA avec variables explicatives. Nous donnons enfin des exemples de modèle explicatif à forme non linéaire.

8.3.2.1. Régression vs ARIMA

A titre d'exemple de modèle de décomposition à tendance déterministe avec variables explicatives, nous citerons Scott (1986) qui utilise dans un premier temps une structure ARIMA après régression préalable du nombre mensuel d'accidents en Grande Bretagne de 1970 à 1978 sur des variables exogènes mesurant le trafic, le prix des carburants, la température, la hauteur de pluie et le nombre de jours ouvrables (ou encore régression avec résidus ARIMA) ; il montre ensuite que la structure ARIMA sur les résidus de la régression peut être omise sous réserve de prendre en compte dans la régression la tendance et la saisonnalité, sous la forme d'une variable temporelle et d'indicatrices saisonnières :

$$\log Y_t = a + bt + S_t + \sum \beta_i \text{Log} X_{it} + \sum_j \beta_j X_{jt} + \lambda \omega_{1t} + \lambda_2 \omega_{2t} + u_t$$

avec : Y le nombre mensuel d'accidents en Grande Bretagne,

$a + bt$ la tendance,

S_t la saisonnalité, mesurée avec onze variables indicatrices,

$X_i, i = 1, 2$: le trafic pour deux catégories de véhicules et le prix des carburants,

$X_j, j = 1 \text{ à } 3$: les deux variables météorologiques et le nombre de jours ouvrables,

ω_{1t} et ω_{2t} deux indicatrices du choc pétrolier de 1974 et de la limitation de vitesse en rase campagne.

8.3.2.2. Modèles structurels

Le modèle structurel de Harvey avec variables explicatives et interventions constitue une forme plus générale d'un modèle de décomposition stochastique que le modèle structurel de base. Utilisé sur des données d'accidents en Grande-Bretagne, il inclut deux variables explicatives (le prix des carburants et l'indice de trafic) qui ont un effet sur la composante tendancielle de Y_t , et il en est de même pour la variable indicatrice $\omega_{t=1_{\geq \tau}}$ qui est utilisée pour évaluer l'impact $\lambda\omega_t$ de la loi sur le port de ceinture.

$$\left\{ \begin{array}{l} \text{Log} Y_t = \mu_t + \gamma_t + \sum_{i=1}^I \beta_i \text{Log} x_{it} + \lambda\omega_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t = \beta_{t-1} + \zeta_t \\ \gamma_t = \sum_{j=1}^{s/2} \gamma_{jt} \\ \gamma_{jt} = (\cos \frac{2\pi j}{s}) \gamma_{j,t-1} + \omega_{jt} \end{array} \right.$$

avec : Y_t le nombre mensuel de conducteurs, tués et blessés graves,

β la pente de la tendance μ , et γ la composante saisonnière,

ε , η , ζ et ω_{jt} des bruits blancs de variances σ_ε^2 , σ_η^2 , σ_ζ^2 et σ_ω^2 , non corrélés entre eux.

De manière équivalente sur données annuelles, la formulation la plus large proposée par Lassarre pour le modèle à tendance linéaire locale incorpore des indicatrices annuelles d'intervention ω_{it} , ω_{jt} et ω_{kt} qui peuvent modifier la composante irrégulière, le niveau ou la pente de la tendance dans le modèle suivant :

$$\begin{aligned} \text{Log}F_t &= \eta \text{Log}V_t + \mu_t + \sum_i \lambda_i \omega_{it} + \varepsilon_t \\ \mu_t &= \mu_{t-1} + b_{t-1} + \sum_j \lambda_j \omega_{jt} + \eta_t \\ \beta_t &= \beta_{t-1} + \sum_k \lambda_k \omega_{kt} + \xi_t \end{aligned}$$

L'utilisation de ce modèle sur les données agrégées de tués de plusieurs pays européens a permis de proposer une évaluation de l'effet des grandes mesures de sécurité routière. Pour la France, ce sont les mesures prises en 1973 - limitation de la vitesse et obligation du taux de port de la ceinture – qui font baisser de manière significative le taux de tués de 17% à compter de 1973 ; une baisse de 9,3% en 1978 est due à l'introduction des contrôles aléatoires d'alcoolémie sur route (Lassarre, 2001).

8.3.2.3. Modèles ARIMA

Les modèles ARIMA avec variables explicatives ont été très fréquemment utilisés sur données mensuelles, principalement pour l'évaluation des effets des mesures de sécurité routière. Ils intègrent généralement les effets exogènes reconnus, c'est-à-dire celui de l'exposition au risque, l'influence du climat par une ou deux variables météorologiques, et l'impact de la configuration calendaire. On utilise des fonctions de transfert lorsque les variables explicatives sont aléatoires, et des modèles d'interventions pour évaluer les effets des mesures de sécurité routière.

Nous citerons à titre d'exemple les modèles développés sur données agrégées espagnoles et françaises.

Ce sont deux variables de consommation de carburants (la consommation d'essence et gazole) à la place du trafic, une variable tenant compte du nombre de week-ends dans le mois WEND et une variable d'intervention représentative d'un grand nombre de mesures de sécurité routière mises en place à partir de juin 1992 $LS^{6/92}$, qui sont utilisées pour modéliser le nombre d'accidents survenus en Espagne de 1982 à 1996 (Rebello, Rivelott, Inglada Lopez de Sabando, dans : COST 329, 2004) :

$$\begin{aligned} \log Y_t &= \sum_i \beta_i \text{Log}X_{it} + \eta \text{WEND} + \gamma LS_t^{6/92} + N_t \\ \nabla \nabla_{12} N_t &= (1 - \theta_1 B)(1 - \theta_2 B^{12}) \varepsilon_t \end{aligned}$$

avec : Y_t le nombre mensuel d'accidents,
 $\nabla\nabla_{12}$ le filtre composé de différences d'ordre 1 et d'ordre 12,
 N_t un premier résidu, non stationnaire,
et ε_t un bruit blanc.

Sur données françaises, la même forme de modèle est utilisée sur données mensuelles d'accidents et de tués. Tout en intégrant l'effet du parcours, de la vitesse, ils permettent surtout d'évaluer l'impact des grandes mesures de sécurité routière. C'est le cas de la première limitation de vitesse de 1973, du choc pétrolier de 1974, de la loi de 1978 sur les contrôles aléatoires d'alcoolémie (Lassarre, Tan dans : Lassarre, 1994).

D'autres modèles du même type, explicatifs des nombres mensuels d'accidents et de tués, ont été développés par grande catégorie de réseaux : routes nationales et autoroutes, réseau départemental, milieu urbain, avec des variables climatiques⁴⁸ et calendaires pour prendre en compte les facteurs transitoires (Vizatelle, 1990).

8.3.2.4. Modèles non linéaires

Comme nous venons de le voir, des modèles non linéaires ont souvent été transformés en modèles linéaires, par application d'une transformation logarithmique à quelques-unes des variables, dépendante ou indépendantes. Ceci facilite l'estimation du modèle.

La relation multiplicative entre l'exposition au risque et le nombre d'accidents, et entre l'exposition au risque et le nombre de décès, est généralement acceptée. Rappelons simplement ici que le premier exemple de modèle agrégé, proposé par Smeed, relie le nombre d'accidents de la route au nombre de véhicules motorisés et à la population correspondante (soit D, M et P respectivement) de manière multiplicative :

$$D = c(MP^2)^{1/3}$$

⁴⁸ S'agissant du facteur climatique dont la mesure nécessite l'accès à des bases de données météorologiques relativement désagrégées, dans le temps et dans l'espace, on notera la contribution de Fridstrom qui pu prendre en compte ce facteur de risque en mesurant les effets de la luminosité et d'un grand nombre de variables climatiques sur des données d'accidents pour quatre pays nordiques Danemark-Finlande-Norvège-Suède (Fridstrom et al, 1995), puis sur une base complète d'indicateurs de risque en Norvège en y appliquant l'approche DRAG (Fridstrom, dans : Gaudry, Lassarre (Eds.), 2000). On notera que le cadre des modèles gaussiens, auquel nous nous limitons ici, n'est pas adapté au cas des petits effectifs, Fridstrom formulant de préférence une hypothèse de régression poissonnienne dans ses travaux.

D'autres transformations peuvent être préférées à la transformation logarithmique, et appliquées aux données observées. Citons le modèle explicatif à trois niveaux, à fréquence mensuelle, proposé par Gaudry, qui repose encore sur une structure de régression multiple avec erreurs autocorrélées et hétéroscédastiques, et avec prise en compte d'une forme de non linéarité. La prise en compte d'un grand nombre de variables explicatives permet de modéliser la tendance et la saison, qui ne sont donc pas filtrées. L'introduction de la transformation de Box-Cox permet une formulation plus souple (forme linéaire, logarithmique ou un compromis) du lien entre la variable endogène et chaque variable exogène.

Le modèle générique s'écrit de la manière suivante, dans laquelle le paramètre $\lambda=(\lambda_Y, \lambda_{X_1}, \dots, \lambda_{X_K})$ est estimé simultanément aux autres paramètres :

$$\begin{cases} Y_t^{(\lambda_Y)} &= \sum_{k=1}^K \beta_k X_{kt}^{(\lambda_{X_k})} + u_t \\ u_t &= v_t \sqrt{\exp\left(\sum_{m=1}^M \delta_m Z_{mt}^{(\lambda_{Z_m})}\right)} \\ v_t &= \sum_{l=1}^p \rho_l v_{t-l} + w_t \end{cases}$$

avec: Y_t la variable endogène à modéliser,

X_{kt} , $k=1$ à K , les variables exogènes (ou explicatives),

u_t le premier résidu, v_t le résidu final,

et w_t un bruit blanc.

et la transformation de Box-Cox définie comme une fonction puissance, de paramètre λ , de toute variable réelle positive V_t par:

$$\begin{aligned} V_t \rightarrow V_t^{(\lambda)} &= \frac{V_t^\lambda - 1}{\lambda} \text{ si } \lambda \neq 0 \\ V_t^{(0)} &= \text{Log } V_t \end{aligned}$$

Dans cette formulation générale, les paramètres de Box-Cox $\lambda_Y, \lambda_{X_1}, \dots, \lambda_{X_K}$ sont estimés en plus des autres paramètres β_k, δ_m et ρ_l , pour $k=1$ à K , $m=1$ à M et $l=1$ à L .

8.3.2.5. Résumé

En résumé, différents types et différentes classes de modèles de séries temporelles ont été retenus pour modéliser des indicateurs agrégés de risque, au niveau national en Europe. Une différence est l'appel à un plus ou moins grand nombre de variables exogènes, mais un autre trait important tient à leur nature, qui peut être déterministe ou stochastique.

8.4. Les facteurs de nature transitoire

Nous nous intéressons plus particulièrement aux facteurs conjoncturels de risque routier, de nature climatique et calendaire.

Comme nous l'avons vu, la plupart des modèles mensuels agrégés de risque routier⁴⁹ les prennent en compte comme facteurs de risque, sous la forme minimale d'un petit nombre de variables relatives à la température, à la pluie, et à une typologie de jours de l'année. Mais ils interviennent également de manière indirecte comme facteurs de l'exposition au risque - communément mesurée par le trafic.

Ces facteurs conjoncturels occasionnent des fluctuations journalières très importantes du volume de trafic, et des nombres d'accidents corporels et de victimes. Leur ampleur est plus marquée au niveau du jour, mais reste non négligeable au niveau du mois.

Le choix des variables associées à ces facteurs conjoncturels, que nous abordons maintenant, dépend d'abord de la nature de l'effet à modéliser : effet périodique ou effet non périodique ? effet global ou effet décomposé ? et dans le cas d'un effet décomposé, selon quel type de décomposition ? et à quel niveau de désagrégation ?

Pour le facteur climatique, il s'agit de sélectionner des variables associées à un phénomène météorologique, qui peuvent être quantitatives ou qualitatives. Quels phénomènes climatiques sont à retenir ? Faut-il retenir un grand nombre de variables météorologiques décrivant ces phénomènes, de manière à évaluer un effet global ? Faut-il choisir un petit nombre de variables météorologiques relatives aux principaux phénomènes d'intérêt de manière à privilégier l'interprétation de leurs paramètres ? Enfin, comment mesurer un phénomène climatique afin qu'il soit représentatif sur l'ensemble d'un territoire géographique ?

⁴⁹ Comme indiqué en 8.2, le risque routier a deux dimensions : le risque d'accident et risque d'être victime dans un accident de la route, à niveau de trafic donné.

Pour le facteur calendaire, il s'agit de construire des indicatrices rendant compte de l'existence de certains jours de l'année, témoins d'un effet calendaire. Comment identifier ces jours ? Quelle typologie de jours retenir ? Puis, comment résumer cette typologie par passage du niveau du jour au niveau du mois ?

8.4.1. Le facteur climatique

Lorsque l'on cherche à évaluer, au niveau d'un pays ou d'une zone géographique, l'influence des facteurs climatiques sur les indicateurs de l'insécurité routière que sont les nombres d'accidents corporels et de victimes de la circulation routière, on distingue naturellement deux groupes de facteurs climatiques. D'une part *ceux qui influencent la mobilité*, et donc de manière indirecte le niveau d'insécurité routière par le bais du trafic, indicateur d'exposition au risque et principal facteur de risque; et d'autre part *ceux qui ont un effet sur le risque routier* : risque d'accident et risque d'être victime (tué, blessé grave, blessé léger) à niveau de trafic donné.

Certains facteurs climatiques, comme des températures clémentes ou la luminosité dans certains pays, favorisent la mobilité ; d'autres, comme les précipitations (pluie, neige), ou comme un froid prolongé, la minorent.

L'influence des facteurs climatiques sur la mobilité agrégée a été étudiée sur le réseau routier national (routes nationales et autoroutes), sur lequel le trafic est mesuré avec une périodicité fine, comme nous l'avons vu dans le chapitre 6. La température, fortement liée au kilométrage parcouru, explique bien l'évolution de la mobilité au cours de l'année. Pour évaluer les effets de situations de froid prolongé, pendant lesquelles la circulation routière peut être sensiblement réduite, il faut faire appel à des facteurs supplémentaires, tels que le gel. Enfin la pluie, qui réduit la mobilité ou la reporte, intervient également indépendamment des facteurs météorologiques précédemment cités.

Les facteurs climatiques qui ont un effet sur le risque routier sont connus. En terme d'enjeu (nombres d'accidents concernés), la pluie occupe la première place, bien avant le brouillard et les autres précipitations : neige/grêle. Les moyennes enregistrées sur la période 1990-2000 sont de 14 % d'accidents corporels par temps de pluie, et de l'ordre de 1% seulement par temps de brouillard, et par temps de neige/grêle (La sécurité routière en France : bilan annuel de l'année, ONISR, 2006).

Ainsi, deux principaux facteurs climatiques peuvent être retenus : la pluie pour ce qui concerne le risque, et la température pour ce qui concerne l'exposition au risque. Le gel (température négative) qui permet de prendre en compte le froid prolongé, et qui intervient dans une moindre mesure dans les deux cas, peut être également retenu. Cette option a été retenue dans le chapitre 10. La question du lien existant entre ces trois facteurs - et plus particulièrement entre les variables qui vont être utilisées pour les prendre en compte dans un modèle explicatif des indicateurs de risque - ne pose pas problème a priori, sauf dans quelques situations extrêmes : canicule, sécheresse (une température élevée est associée à une absence de pluie) et gel (une température faible est associée au gel), qui sont rares et qui n'ont pas donné lieu à un traitement particulier.

Une autre manière de procéder consiste à retenir un spectre très large de facteurs climatiques. Cette option conduit à évaluer un effet météorologique d'ensemble, global, tous facteurs climatiques confondus. L'interprétation des paramètres relatifs aux variables climatiques est à exclure, mais en revanche l'amplitude de l'effet météorologique est mieux évaluée. Cette option a été retenue dans le chapitre 9.

8.4.2. Le facteur calendaire

Les conséquences de la configuration du calendrier sur le niveau de l'insécurité routière sont médiatisées : les fins de semaine sont des périodes accidentogènes, certains week-ends sont connus pour être particulièrement meurtriers, en particulier lorsqu'ils se situent au voisinage d'un jour férié ou d'une date de départ ou de retour de congés. Là encore, l'effet indirect de la configuration calendaire sur les nombres d'accidents et de victimes qui se manifeste par le biais du trafic, est à distinguer de l'effet direct, et pourrait être relativement important. L'effet calendaire direct, d'une variation du nombre des accidents et du nombre de tués à niveau de trafic donné, traduit une variation du risque d'accident et de sa gravité à certains moments de l'année.

Les effets calendaires sont de deux types : *un effet périodique*, qui se reproduit chaque semaine (effet du jour de semaine) ou chaque année (effet du mois de l'année) et *un effet exceptionnel*, discontinu, non périodique, qui se manifeste au voisinage des jours fériés et des périodes de congés. Alors que le calendrier des congés est prédéterminé, la difficulté réside dans la détermination précise de ce voisinage.

L'identification des jours de l'année témoins d'un effet calendaire exceptionnel est plus facile à mener sur les indicateurs de trafic que sur les indicateurs de risque. Les indicateurs de trafic présentent par nature une plus grande stabilité et constituent un indicateur "témoin" d'un effet calendaire. La méthode préconisée a déjà été mentionnée dans le chapitre 6, et sera détaillée plus avant dans le chapitre 9.

8.4.3. Applications retenues

Dans les chapitres qui suivent, nous présentons trois types d'applications qui ont été réalisées en prenant en compte les effets de ces facteurs conjoncturels sur les indicateurs du risque routier.

Il s'agit du modèle Giboulée (chapitre 9) et du modèle RES (chapitre 10), utilisés pour le suivi et la prévision des indicateurs d'insécurité routière. Une approche multivariée du risque tel que conçu dans le modèle RES est également discutée (chapitre 11). Nous présentons enfin une utilisation des modèles Giboulée et RES dans un objectif spécifique, qui a été de déterminer l'existence d'un lien statistique entre l'annonce anticipée d'amnistie des fautes de conduite qui accompagne l'élection présidentielle et l'accidentalité (chapitre 12).

Chapitre 9.

Le modèle Giboulée : un modèle de suivi à court terme

Nous présentons dans ce chapitre des travaux réalisés à la demande de la DSCR, qui ont donné lieu au rapport du groupe de travail OEST/CEMS/INRETS/SETRA (Bergel et al, 1995a). Ils s'appuient sur une approche méthodologique réalisée en collaboration avec une équipe des universités de Paris-1 et de Paris-X12 (Azencott et al, 1991 et 1992) D'autres présentations en sont faites dans (Bergel, Girard, 1993), et (Bergel et al, 1994, 1995b et 1995c).

9.1. Objectif

La première application porte sur la construction d'un tableau de bord mensuel de l'insécurité routière, permettant une analyse au niveau du jour.

La demande de la DSCR portait sur un suivi des nombres mensuels d'accidents corporels et de victimes (tués, blessés graves, blessés légers) agrégés au niveau France entière, qui fasse dans la mesure du possible la part des facteurs déterminants de l'insécurité routière, et qui permette de dégager une tendance.

9.2. Méthode

La démarche retenue a consisté à construire un modèle journalier, complétant un modèle journalier de régression linéaire déjà développé au SETRA (Le Breton, Vervialle, 1990), qui reproduise au mieux la dynamique observée en sécurité routière et à en tirer un modèle mensuel. Des statistiques journalières étaient disponibles sur une période de six ans, avec une désagrégation par grande catégorie de réseau.

Les différences marquées, de niveau et d'évolution, du risque routier par réseau, et le fait que les statistiques d'accidents de victimes étaient disponibles selon ce même découpage ont conduit à une modélisation par type de réseau (routes nationales, autoroutes, réseau secondaire et milieu urbain).

Parmi les facteurs de risque, les conditions de circulation (conditions météorologiques, niveaux de trafic et d'encombrement) qui peuvent varier très fortement d'un jour à l'autre

constituent le principal déterminant de la variabilité des statistiques journalières d'accidents et de victimes.

Cependant l'approche était contrainte par la disponibilité de données journalières relatives à des facteurs de risque. Or le fait que le volume de trafic journalier, premier facteur de risque, ne soit pas mesuré sur tous les réseaux a conduit à ne pas utiliser cette mesure partielle de l'exposition au risque, que ce soit dans les modèles de nombres d'accidents ou de nombres de victimes⁵⁰.

En revanche, les données relatives aux conditions météorologiques et à la configuration du calendrier, qui sont disponibles en rythme journalier et pour l'ensemble du territoire, ont été retenues pour modéliser l'effet des conditions de circulation sur les indicateurs journaliers.

9.2.1. Le modèle

Au niveau du jour, un modèle de décomposition additive a été retenu, dont les composantes peuvent être qualifiées d'“effets” (cf. figure 9.1). La décomposition est la suivante :

$$Y_t - YM_t - YC_t = T_t + H_t + u_t,$$

avec : Y_t la variable à modéliser, et t le jour,

YM_t un effet météorologique de court terme, qui agit instantanément, au jour le jour; mesuré par différence avec un effet de normale saisonnière - on peut aussi parler d'un effet météorologique local, dans le temps,

YC_t un effet calendaire exceptionnel au voisinage des jours fériés, au moment des départs et des retours de congés, et pendant les vacances scolaires,

H_t un effet calendaire hebdomadaire régulier⁵¹ qui prend en compte les variations journalières régulières à l'intérieur de la semaine,

T_t une tendance⁵² courte, calculée sur plusieurs jours, sur laquelle les modifications rapides liées à des facteurs non modélisés (par exemple l'entrée en vigueur d'une mesure de sécurité routière) peuvent être détectées,

⁵⁰ Même sur le réseau national (routes et autoroutes) sur lequel il est mesuré, le niveau de trafic peut ne pas suffire à modéliser les variations importantes des nombres d'accidents et de victimes qu'on observe certains jours de l'année. On est donc naturellement amené à construire des variables calendaires qui rendent compte globalement des conditions de circulation, sans isoler le trafic.

et u_t un résidu, stationnaire, non corrélé avec T_t , H_t et YC_t .

Les composantes journalières sont ensuite agrégées mensuellement (cf. figure 9.2).

Au niveau du mois, trois composantes sont constituées par agrégation des composantes journalières :

$$YM_m = \sum_{t \in m} YM_t, YC_m = \sum_{t \in m} YC_t \text{ et } H_m = \sum_{t \in m} H_t$$

Notons : $Ycvl_t = Y_t - YM_t - YC_t - H_t = T_t + u_t$,

$$Ycvl_m = Y_m - YM_m - YC_m - H_m,$$

avec : $Y_m = \sum_{t \in m} Y_t$

La tendance courte, agrégée mensuellement, se décompose en deux parties :

$$\sum_{t \in m} T_t = S_m + T_m, \text{ pour chaque mois } m,$$

avec : S_m un effet mensuel global, qui prend en compte les grandes variations à l'intérieur de l'année⁵³,

et T_m une tendance longue, calculée sur plusieurs mois.

La décomposition mensuelle complète est ainsi :

$$Y_m = YM_m + YC_m + H_m + S_m + T_m + u_m$$

⁵¹ c'est-à-dire que la différence $H_t - H_{t-7}$ est stationnaire

⁵² c'est-à-dire que la différence $T_t - T_{t-1}$ est stationnaire

⁵³ mais pas les variations régulières à l'intérieur de la semaine, l'effet hebdomadaire ayant été extrait au niveau du jour

9.2.2. Options retenues

Des variables météorologiques et des variables calendaires sont utilisées pour modéliser les deux premières composantes YM_t et YC_t . Des données de ces deux types avaient déjà été utilisées pour le modèle journalier existant, mais la manière de les constituer a été repensée, afin qu'elle soient adaptées à la nouvelle forme de modèle retenue.

En revanche, la composante hebdomadaire et la tendance courte, H_t et T_t , sont estimées indirectement par des techniques de lissage. De la même manière, sur la série corrigée des effets locaux (météorologique, calendaire exceptionnel et calendaire hebdomadaire) et agrégée mensuellement, les composantes S_m et T_m sont obtenues par lissage.

Les choix possibles pour la modélisation des deux premières composantes, à l'aide de variables exogènes, portaient :

- sur le caractère séquentiel : modéliser en simultané ou de manière indépendante l'effet météorologique et l'effet calendaire ?
- et sur le type de modèle : modèle de régression linéaire ou modèle ARIMA ?

La quantification des deux effets météorologique et calendaire exceptionnel a été réalisée en deux étapes successives, en supposant que l'influence locale du climat (c'est à dire de la différence à la normale saisonnière) est indépendante de la configuration calendaire (calendrier exceptionnel et calendrier hebdomadaire). De fait, les variables météorologiques, mesurées par différence avec des normales saisonnières, ne sont pas corrélées avec les indicatrices journalières qui codent la présence des jours de l'année témoins d'un effet calendaire exceptionnel.

Cette manière séquentielle de procéder permet d'utiliser des techniques statistiques différentes pour chaque séquence : modèle de régression linéaire avec stepwise sur un très grand nombre de variables météorologiques pour les données brutes, puis modèle ARIMA avec un petit nombre de variables calendaires exogènes pour les données corrigées de l'effet météorologique de court terme (l'effet calendaire exceptionnel qui s'observe souvent sur quelques jours consécutifs est ainsi pris en compte par une modélisation dynamique).

9.2.3. Innovation méthodologique

La double correction de ces deux effets météorologique et calendaire constitue une correction locale originale, absente de la méthode standard de correction des variations saisonnières, et adaptée au domaine de la sécurité routière dans la mesure où les paramètres exogènes sont estimés par indicateur d'insécurité routière. Notons à ce stade que cette méthode est extrapolable à tout autre domaine d'application pour lequel des statistiques journalières sont disponibles sur une assez longue période.

On obtient au niveau du jour une série corrigée des variations locales ou cvl qui est encore, après agrégation mensuelle, désaisonnalisée par une méthode classique telle que la succession de lissages mise en œuvre par le programme Census-X11. Au niveau du mois, les composantes sont au nombre de six, dont les trois premières sont estimées sur une base journalière.

Nous nous intéressons maintenant exclusivement à la partie du modèle journalier pour laquelle des variables exogènes sont utilisées, c'est à dire la modélisation des effets météorologique et calendaire exceptionnel YM_t et YC_t (Azencott et al., 1991 et 1992).

L'originalité de la composante météorologique tient au fait qu'elle a été établie à partir d'une information régionale, mesurée en 6 stations météorologiques (ou stations pilotes : Bordeaux, Marignane, Grenoble, Nancy, Orléans et Dinard), avec des variables explicatives de l'influence du climat sur la fréquence et la gravité des accidents (ou clignotants). Ces clignotants diffèrent en effet selon l'indicateur modélisé, et selon la saison.

La composante calendaire exceptionnelle est estimée par appel à des variables auxiliaires (variables calendaires), codant des types de jours "exceptionnels" dans l'année. Bien que ces jours soient identifiés sur des indicateurs de trafic, l'évaluation de l'effet calendaire exceptionnel est ensuite réalisée par indicateur d'insécurité routière.

Nous exposons maintenant de manière plus détaillée la manière d'estimer les deux composantes d'intérêt.

9.2.3.1. L'estimation de l'effet météorologique de court terme YM_t

Le volume des données existantes fournies par la Météorologie Nationale - données quantitatives journalières, parfois même disponibles par tranche horaire, observées en une centaine de stations réparties sur le territoire national - nécessitait de résumer cette information. Il était nécessaire de faire un tri parmi la cinquantaine de variables fournies : une douzaine de variables ont été sélectionnées, pour représenter une situation climatique génératrice d'insécurité.

L'on a également cherché à disposer d'une mesure plus fine qu'une simple moyenne nationale de données observées localement, qui soit représentative des disparités climatiques régionales. Six régions homogènes au plan météorologique ont été constituées, et une station représentative de chaque région a été retenue. On dispose ainsi de 72 variables météorologiques pour estimer l'effet météorologique. Au final, 15 à 30 variables, selon l'indicateur et selon la saison, sont retenues à l'issue de régressions linéaires avec stepwise de chaque indicateur sur ces 72 variables, à un seuil de 30%.

9.2.3.2. L'estimation de l'effet calendaire exceptionnel YC_t

Les effets calendaires sur les nombres journaliers d'accidents et de victimes sont identifiables, à l'examen des graphiques journaliers de ces indicateurs. Mais leur étude qualitative (présence/absence de l'effet) est simplifiée si on la mène sur les séries de trafic, qui présentent par nature une plus grande stabilité, et qui constituent un indicateur "témoin" permettant de déceler la présence d'un effet calendaire. L'évaluation de ces effets, une fois identifiés, est ensuite réalisée par indicateur d'insécurité routière. On suppose en quelque sorte qu'une variation importante du niveau de trafic modifie les conditions de circulation et de conduite, et de façon différente selon la nature du déplacement; le nombre et la gravité des accidents peuvent être affectés par l'ensemble de ces phénomènes.

Les variables auxiliaires créées pour modéliser l'effet calendaire exceptionnel se classent en trois groupes, selon qu'elles signalent un jour de vacances scolaires, un jour affecté par un départ ou un retour de congés, ou par la proximité d'un jour férié.

Une fois ces variables calendaires constituées, l'influence de la configuration du calendrier est évaluée par indicateur d'insécurité routière, et sur la série journalière corrigée de l'effet météorologique de court terme, issue des régressions précédentes. C'est la modélisation

ARIMA avec variables exogènes qui a été retenue, en raison de ses propriétés adaptatives: modélisation de la dynamique de la série, par la prise en compte de ses autocorrélations. L'effet calendaire hebdomadaire est ensuite calculé à partir de l'indicateur corrigé des effets météorologique local et calendaire exceptionnel.

9.3. Les principaux résultats

On se limitera ici à l'exposé des principaux résultats, obtenus pour le nombre de tués France entière, de février 1985 à décembre 1992.

9.3.1. Les sorties journalières

On dispose d'abord des deux séries journalières d'effets locaux: un effet météorologique et un effet calendaire sur la période des huit ans, de février 1985 à décembre 1992. La série corrigée des effets locaux diffère sensiblement de la série brute, avec des écarts qui peuvent être très importants. Ainsi, les fluctuations du nombre journalier de tués, qui varie entre 10 à 15 en milieu de semaine à 40 voire 45 en fin de semaine au cours du deuxième quadrimestre 1992 sont sensiblement réduites après correction des effets locaux; le seul effet calendaire dépasse régulièrement les 10 tués par jour en fin de semaine, et l'effet météorologique, plus limité en amplitude mais plus irrégulier, peut occasionner 5 voire même 8 tués certains jours (cf. figures 9.3 et 9.4).

La détection d'effets supplémentaires, non pris en compte dans le modèle, peut se faire à l'examen de la tendance. A titre d'exemple, l'instauration du permis à points le 1er juillet 1992, a eu pour effet d'infléchir sensiblement la tendance journalière du nombre de tués, alors que le profil moyen est à la progression à cette période de l'année, comme on l'observe sur les tendances journalières des sept années précédentes (cf. figure 9.5). Une première inflexion à la baisse correspond à la campagne de communication lancée début mai; une seconde inflexion rend compte des effets de l'instauration de la mesure le 1er juillet, qui n'ont été véritablement sensibles qu'à partir de la mi-juillet, en raison des barrages des transporteurs routiers qui ont bloqué du 1er au 9 juillet le réseau routier et autoroutier, occasionnant à cette période de départs en vacances un report important de trafic sur le réseau secondaire, sur lequel la gravité des accidents est plus importante.

9.3.2. Les sorties mensuelles

En rythme mensuel, les effets locaux sont atténués - les effets journaliers positifs et négatifs vont se compenser dans l'agrégation mensuelle. La série mensuelle corrigée des variations

locales diffère encore de la série mensuelle brute, mais leur écart mensuel est réduit en moyenne, même s'il reste très important certains mois. On enregistre sur le nombre de tués France entière, sur la période 1985-1992, des écarts qui peuvent dépasser les 10%, comme en février 86 et janvier 1987, en raison de deux vagues de froid prolongé, puis en juillet 88 par exemple, en raison de la configuration du calendrier de ce mois, qui comptait 5 week-ends: le cumul des deux effets locaux atteint 4, puis même 7 tués par jour en février 1986 et janvier 1987, et à nouveau 4 tués par jour en juillet 1988 (cf. figure 9.6).

9.4. Conclusions et perspectives

En tant que modèle de suivi, ce modèle permet d'abord un suivi conjoncturel plus immédiat et moins biaisé que ne le permettent les techniques de calcul habituellement utilisées sur les données brutes (ratios de comparaison des données journalières ou mensuelles à la même période de l'année précédente, ou lissage par moyenne mobile des données mensuelles sur les douze derniers mois par exemple). Le suivi peut être réalisé d'un jour à l'autre, ou d'un mois au mois suivant; en particulier, les inflexions de la tendance de moyen terme sont saisies dès qu'elles se manifestent, environ six mois avant qu'elles n'apparaissent sur les moyennes mensuelles glissantes des douze derniers mois.

En tant que modèle de désaisonnalisation, ce modèle est à la fois un modèle de désaisonnalisation des indicateurs journaliers - c'est d'abord la saisonnalité irrégulière, due à une situation climatique inhabituelle par rapport à la normale saisonnière, à la configuration du calendrier, qui est évaluée puis corrigée en rythme journalier - et un modèle de désaisonnalisation des indicateurs mensuels - c'est de plus la saisonnalité régulière qui est évaluée puis corrigée en rythme mensuel (cf. figures 9.7 et 9.8). A la différence de la méthode classique de désaisonnalisation de données mensuelles qui portent sur des mois non homogènes, cette méthode de désaisonnalisation est plus efficace mais elle demande une bonne connaissance des données météorologiques et des événements calendaires.

Ce modèle est enfin approprié au contexte de la sécurité routière, dans lequel le facteur aléatoire est particulièrement important. Cette distinction entre saisonnalité irrégulière et saisonnalité permanente permet que soit quantifiée la partie de l'aléa engendrée par une météorologie inhabituelle et par la configuration du calendrier; il reste néanmoins difficile d'envisager une prise sur cet aléa saisonnier.

Le modèle de suivi journalier est actuellement appliqué aux nombres d'accidents et de victimes de la circulation routière désagrégés par catégories de réseaux et suivant une typologie d'accidents.

Une modélisation directe des indicateurs mensuels sur plus longue période a été également menée. Il s'agit du modèle RES (pour réseaux), agrégé par grande catégories de réseaux, pour l'analyse à court/moyen terme, qui est discuté ci-après dans le chapitre 10.

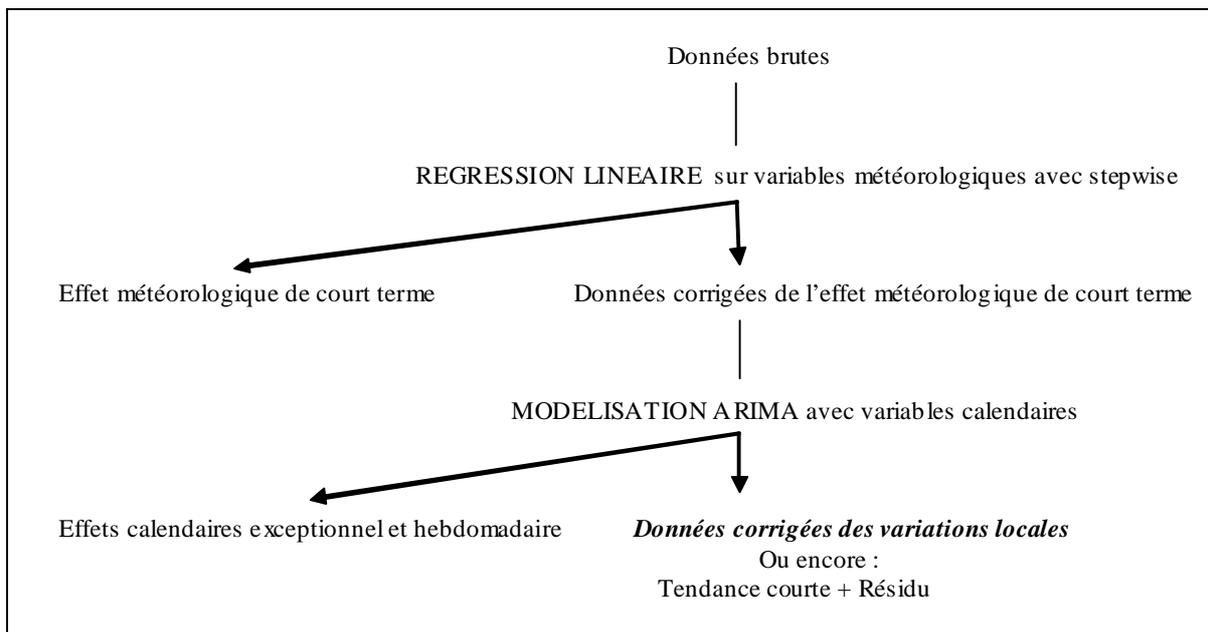


Figure 9.1 La modélisation journalière dans le modèle Giboulée

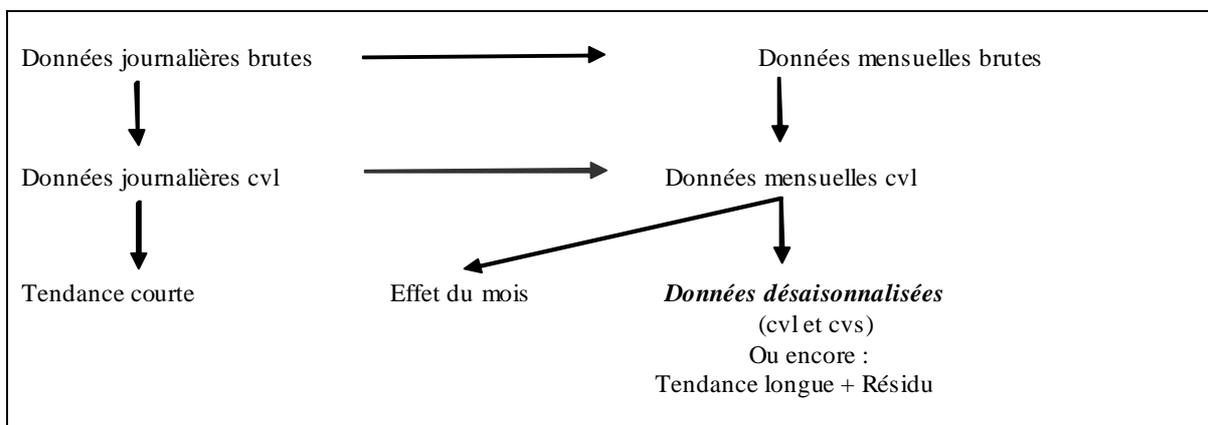


Figure 9.2. La modélisation mensuelle dans le modèle Giboulée

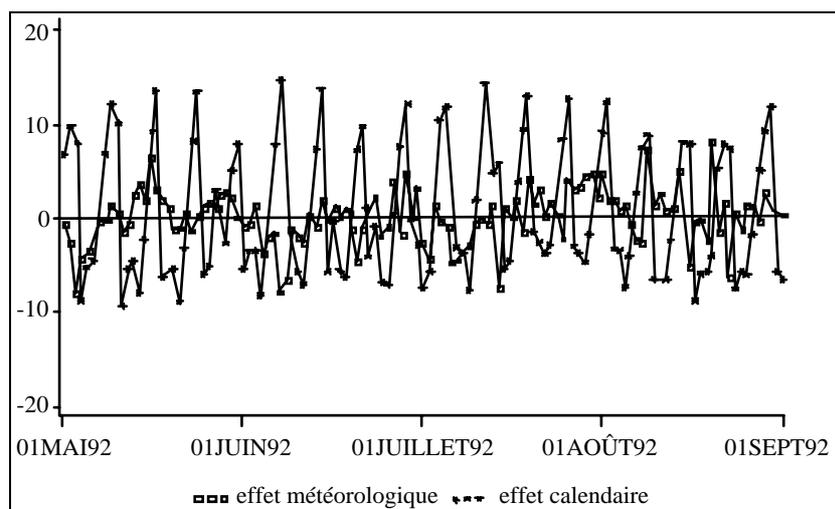


Figure 9.3- Nombre journalier de tués France entière : effets météorologique et calendaire, du 1er mai au 31 août 1992

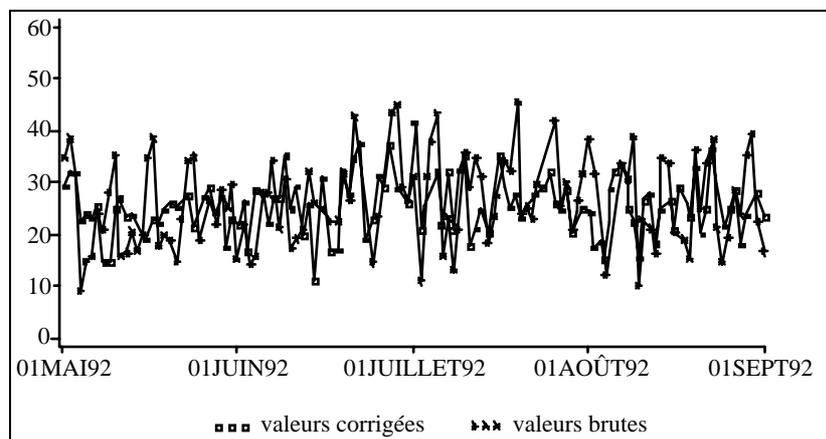


Figure 9.4- Nombre journalier de tués France entière : données brutes et données corrigées des variations locales, du 1er mai au 31 août 1992.

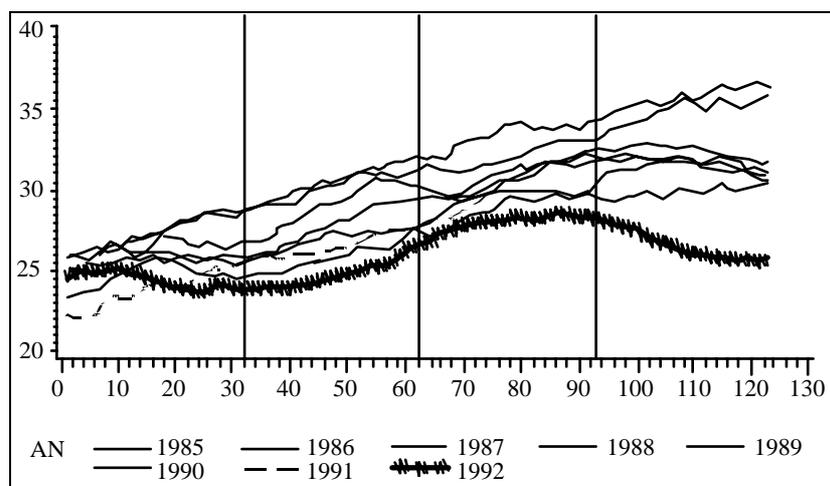


Figure 9.5.- Nombre journalier de tués France entière : évolution de la tendance journalière, du 1er Mai au 31 Août 1992, comparativement à celles des années précédentes.

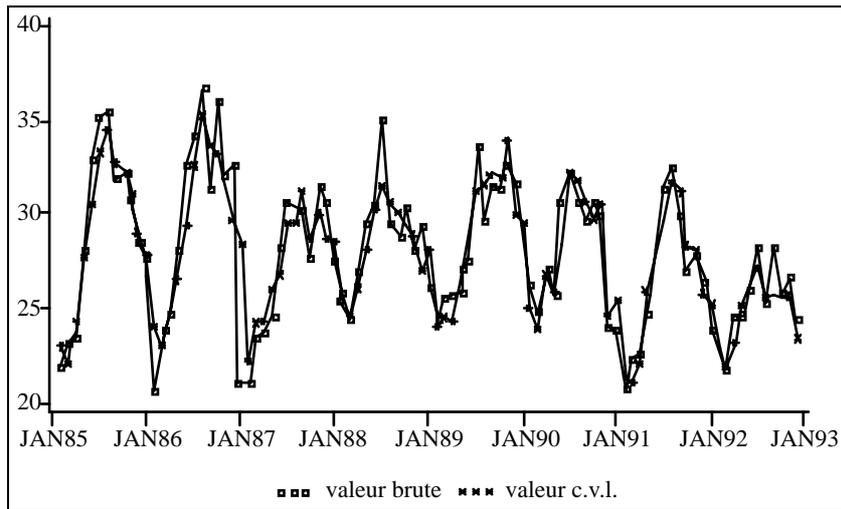


Figure 9.6.- Nombre mensuel de tués France entière : données brutes et données corrigées des variations locales, de février 1985 à décembre 1992.

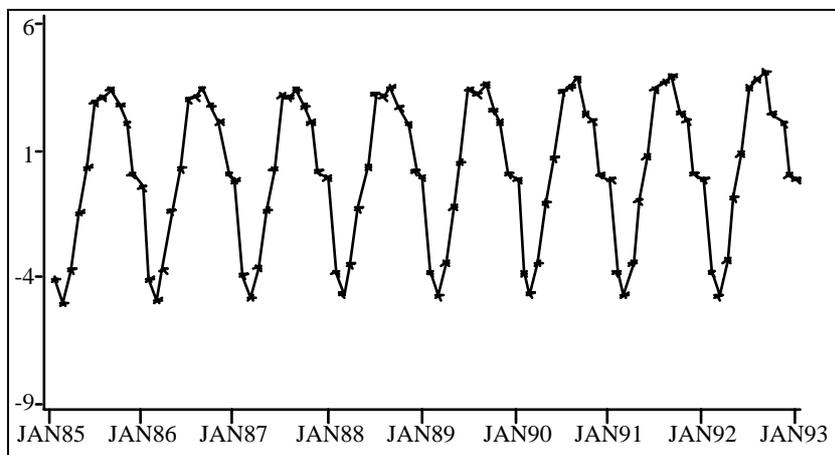


Figure 9.7: Nombre mensuel de tués France entière : effet du mois, de février 1985 à décembre 1992.

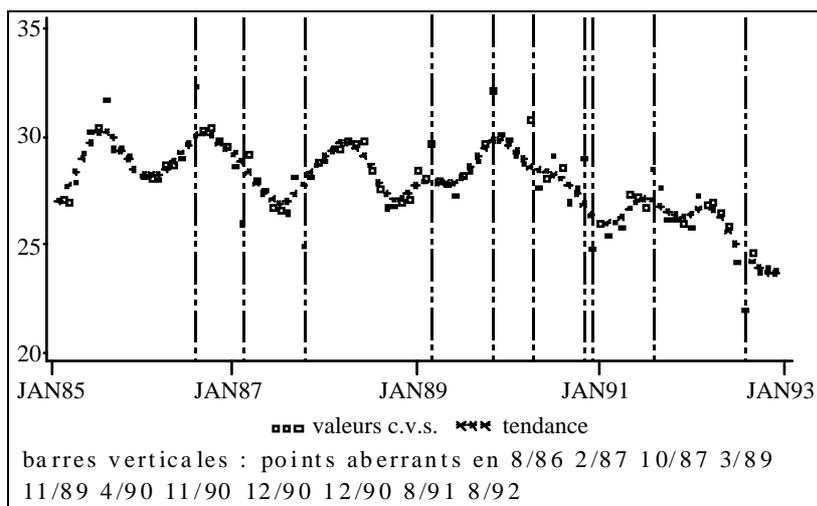


Figure 9.8: Nombre mensuel de tués France entière : données désaisonnalisées et tendance, de février 1985 à décembre 1992.

Chapitre 10.

Le modèle RES : un modèle de court/moyen terme

La présentation qui suit résulte d'un ensemble de travaux réalisés à la demande de la DSCR, dans le cadre de deux conventions passées avec l'Inrets entre 1998 et 2000.

Les références reprises dans ce chapitre sont tirées de deux rapports de convention Inrets/DSCR rendant compte de ces travaux (Bergel, Depire, 2000 et 2002). Des résultats partiels en sont donnés dans (Bergel, Depire, 2004a et b).

10.1. Objectif

Cette deuxième application porte sur la construction d'un modèle permettant de faire de l'analyse de court/moyen terme⁵⁴ de l'évolution des nombres d'accidents et de victimes, agrégés par catégorie de réseau. Il s'est agi de compléter la demande précédente, et d'établir un modèle à petit nombre de facteurs de risque significatifs sur une base mensuelle, qui soit à la fois explicatif et prédictif. Nous nous limitons ici à traiter sa dimension explicative. A la différence de ce qui vient d'être présenté dans le chapitre 9 avec le modèle Giboulée, le pas de temps retenu ici est directement le mois.

10.2. Problématique

Nous avons cherché à nous situer dans un cadre plus large que celui de la régression linéaire multiple, et ainsi d'envisager une forme fonctionnelle plus souple, de type puissance, telle que celle préconisée par Gaudry.

Nous nous référons ici à la forme de modèle autorégressif, explicatif du risque routier, due à Gaudry, exposée dans le chapitre 8, et aux trois niveaux du risque routier de la démarche DRAG : la demande routière, le risque d'accident et la gravité de l'accident, mesurés ici ⁵⁵ par les nombres mensuels de véhicule kilomètres parcourus, d'accidents corporels et de victimes de la circulation routière.

La forme générale de cette approche, donnée en 8.3.2.3, fait appel à un paramètre supplémentaire par variable (le paramètre λ de la transformation de Box-Cox effectuée sur

⁵⁴ soit à un horizon d'un an et plus .

⁵⁵ comme indiqué en 8.2.2.1, les indicateurs qui mesurent le risque d'accident et sa gravité peuvent être des nombres absolus d'accidents et de victimes ou des ratios. C'est la première option qui est retenue ici.

cette variable). L'utiliser pour toutes les variables exogènes nous aurait conduits à doubler le nombre de paramètres à estimer pour une même équation, et cette démarche coûteuse en temps d'estimation et en données ne pouvait pas être retenue. Disposant ici de données mensuelles mesurées sur une vingtaine d'années, nous avons choisi de limiter le nombre de paramètres estimés par équation⁵⁶. Cette contrainte nous conduisait à limiter le nombre de variables exogènes par équation et/ou le nombre de paramètres attribués à une variable.

10.3. Démarche

La démarche a été menée en deux temps : la recherche d'une forme fonctionnelle sur le réseau routier national, et l'extension à d'autres réseaux de cette forme fonctionnelle validée.

Dans un premier temps, nous nous sommes limités à considérer les deux niveaux de risque et de gravité, mesurés par le nombre d'accidents et de tués. La validation de la forme fonctionnelle a d'abord été établie sur le réseau routier national (routes nationales et autoroutes), dans un modèle à deux facteurs de risque : l'exposition au risque mesurée par une variable principale⁵⁷ de trafic, et le facteur climatique mesuré par des variables secondaires relatives à la température, à l'occurrence de gel et à la hauteur de pluie. La transformation de Box-Cox a été retenue pour la variable principale qui mesure l'exposition au risque, et le nombre de paramètres exogènes à estimer a été limité à cinq (pour quatre variables exogènes), ce qui porte à une dizaine le nombre total de paramètres à estimer.

Dans un deuxième temps, nous avons retenu la spécification validée à l'étape précédente, et nous l'avons étendue de la manière suivante. Sur le réseau routier national, nous avons considéré les trois niveaux de la démarche DRAG, mesurés par le nombre de véhicule-kilomètres, le nombre d'accidents et le nombre de tués. Sur deux réseaux additionnels : réseau secondaire, milieu urbain, et sur la France entière, les variables de trafic manquent et ont été remplacées par des variables représentatives de l'offre et de la demande de transport⁵⁸ dans une approche à nouveau limitée à deux niveaux. Lors de cet exercice d'extension, les variables météorologiques, déjà mesurées en moyennes mensuelles, ont été assorties de

⁵⁶ Estimer une dizaine de paramètres par équation revient à disposer de 20 à 30 données par paramètre estimé, selon que la période de calage est 1975-1993 ou 1975-1998

⁵⁷ Nous distinguons ici une variable principale de trafic mesurant l'exposition au risque (l'exposition au risque est considérée comme le premier facteur de risque), et trois variables secondaires mesurant le facteur climatique.

⁵⁸ La consommation de carburants est utilisée comme proxy (mesure) de l'exposition au risque sur l'ensemble du territoire, ce choix a été fait dans le modèle exposé dans le chapitre 12.

variables d'atypicité météorologique témoins de valeurs journalières extrêmes, et de variables calendaires. Dans ce modèle, le nombre de paramètres exogènes à estimer est de seize (pour seize variables exogènes), ce qui porte à une vingtaine le nombre total de paramètres à estimer.

10.4. Recherche de forme fonctionnelle

10.4.1. Objectif

Les deux indicateurs modélisés sont les nombres mensuels d'accidents corporels et de tués, enregistrés sur un réseau donné : routes nationales et autoroutes, parmi lesquelles les autoroutes concédées. Pour chacun de ces indicateurs, les variables explicatives retenues sont d'une part le volume de trafic mensuel enregistré sur le réseau, et d'autre part trois variables météorologiques - la température journalière maximale, l'occurrence journalière de gel et la hauteur journalière de pluie -, agrégées ou moyennées par mois et sur l'ensemble du territoire.

Dans la spécification que nous retenons, décrite dans l'équation 1, la variable endogène subit une transformation logarithmique⁵⁹ et s'exprime comme combinaison linéaire d'une variable de trafic subissant une transformation de Box-Cox de paramètre λ et de variables météorologiques. Cette spécification générale comporte les deux cas particuliers : $\lambda=0$ et $\lambda=1$, correspondant aux équations 2 et 3, et nous chercherons à déterminer si la spécification générale diffère significativement de l'un et de l'autre de ces deux cas particuliers. Dans tous les cas, les variables secondaires ont un effet linéaire sur le logarithme de la variable endogène.

Rappelons la transformation de Box-Cox T_λ de paramètre λ (Box, Cox, 1964) définie par :

$$T_\lambda : \begin{cases} X_t \rightarrow X_t^{(\lambda)} & = \frac{X_t^\lambda - 1}{\lambda} \text{ si } \lambda \neq 0 \\ X_t^{(0)} & = \log X_t \end{cases}$$

et les trois équations considérées :

⁵⁹ Nous ne cherchons pas à appliquer la transformation de Box-Cox à l'endogène Y , car cela nécessite de faire l'hypothèse que pour toute valeur de λ , Y^λ est une variable gaussienne.

Équation 1 : Modèle général

$$\log Y_t = \beta_1 \frac{X_{1,t}^\lambda - 1}{\lambda} + \sum_{j=2}^{J+1} \beta_j X_{j,t} + \mu + \varepsilon_t, \phi(B)\varepsilon_t = u_t$$

Équation 2 : Cas particulier $\lambda=0$

$$\log Y_t = \beta_1 \log X_{1,t} + \sum_{j=2}^{J+1} \beta_j X_{j,t} + \mu + \varepsilon_t, \phi(B)\varepsilon_t = u_t$$

Équation 3 : Cas particulier $\lambda=1$

$$\log Y_t = \beta_1 X_{1,t} + \sum_{j=2}^{J+1} \beta_j X_{j,t} + \mu + \varepsilon_t, \phi(B)\varepsilon_t = u_t$$

où $\phi(B)$ est un polynôme en B représentant la dynamique du processus corrigé des effets exogènes, et u_t un bruit blanc,

avec : pour Y le nombre d'accidents corporels ou de tués sur un réseau donné,
 pour X_1 la variable principale de trafic sur le même réseau,
 pour X_j $j=2, \dots, J+1$ les J variables climatiques secondaires.

Après avoir estimé le modèle de l'équation 1 pour les 6 équations considérées (c'est-à-dire pour chacun des deux indicateurs et pour chacun des 3 réseaux) sur la période 1975-1993, nous réalisons les tests d'hypothèse : $\lambda=0$ puis $\lambda=1$, en comparant la vraisemblance des modèles des équations 1 et 2 (respectivement 1 et 3).

Nous cherchons ensuite à juger de la stabilité des résultats obtenus par extension de la période d'étude à 1975-1998. Enfin, nous interprétons les résultats obtenus.

10.4.2. Méthode

10.4.2.1. Estimation du modèle de l'équation 1

La difficulté principale est l'estimation non linéaire du paramètre de Box-Cox. Les procédures classiques disponibles dans les logiciels courants ne permettent pas d'obtenir facilement une estimation. En effet d'une part certaines estiment des paramètres dans un cadre non linéaire

(procédure NLIN dans SAS) mais ne sont pas adaptées à la modélisation des séries temporelles, d'autres s'emploient dans un cadre temporel (procédure AUTOREG et procédure ARIMA) mais ne permettent pas de paramétrer une transformation des variables exogènes.

Nous avons utilisé l'algorithme d'estimation linéaire, en supposant que pour tout indicateur et sur tout réseau, le logarithme de l'indicateur modélisé, conditionnellement aux variables exogènes, suit une loi approximativement normale.

Dans l'algorithme d'estimation utilisé, le paramètre de Box-Cox va successivement parcourir l'intervalle $[-4 ; +4]$ avec un pas δ . Une estimation par le maximum de vraisemblance du modèle ARMAX dont l'exogène principale est $\frac{X^{\lambda_i} - 1}{\lambda_i}$ est réalisée à chaque pas i . C'est la procédure AUTOREG de SAS, appropriée pour estimer un modèle de régression linéaire avec résidus autocorrélés, qui est utilisée pour mettre en œuvre l'estimation par maximum de vraisemblance.

Une fois la procédure achevée, on retient la valeur du paramètre associée à la plus petite des 100 valeurs prises par la fonction objectif.

Au final, pour le modèle retenu, les tests usuels sur les résidus sont effectués : tests d'indépendance (test des up and down), de bruit blanc (test utilisant la statistique de Shapiro-Wilk), de normalité (test de Fisher) ; Ils sont dans l'ensemble satisfaisants, voir tableau 2a ; en particulier, l'hypothèse de non autocorrélation (test de Ljung-Box) a été validée.

10.4.2.2. Test d'hypothèse

Après estimation du paramètre λ de Box-Cox, l'hypothèse $\lambda=0$ (respectivement $\lambda=1$) est testée en comparant la vraisemblance des deux modèles emboîtés, des équations 1 et 2 (respectivement 1 et 3).

La statistique G utilisée (Enders, 1995) suit asymptotiquement, sous l'hypothèse H_0 ($\lambda=0$ ou $\lambda=1$), une loi de χ^2 centré à d degrés de liberté :

$$G = -2(\log V_r - \log V_u) \xrightarrow{L} \chi_d^2$$

avec : V_u la vraisemblance du modèle où λ a été estimé (unrestricted, ou non contraint),

V_r , la vraisemblance du modèle où λ a été fixé, à 0 ou à 1 (restricted, ou contraint),
et le nombre d de contraintes ici égal à 1.

10.4.2.3. Stabilité

La stabilité des résultats a été étudiée par extension de la période de calage, qui passe de 1975-1993 à 1975-1998, et s'entend de la manière suivante :

stabilité de la valeur du paramètre de Box-Cox correspondant à la variable principale;
stabilité des estimations des paramètres des autres variables,
stabilité en terme de réponse aux tests de comparaison de modèles.

Pour le paramètre de Box-Cox, un intervalle de confiance n'a pas été calculé du fait de la méthode utilisée, et la stabilité ne peut donc pas être étudiée. Pour les paramètres des variables secondaires, et pour ceux de la dynamique, on a deux estimations pour le même paramètre, sur la période initiale et sur la période étendue, et on considère qu'il y a stabilité si l'estimation 1 est incluse dans l'intervalle de confiance à 95% de l'estimation 2, et inversement.

10.4.3. Résultats

Les résultats que nous donnons ici sont résumés dans les Tableaux 10.1 et 10.2.

Les modèles avec forme fonctionnelle optimale (c'est le cas de la spécification initiale retenue ici, dans laquelle le paramètre λ n'est pas contraint) produisent dans tous les cas, sur la période initiale et sur la période étendue, une estimation significative du paramètre relatif à l'exogène principale (β_1), alors que les estimations des paramètres relatifs aux variables secondaires ne le sont pas toujours au seuil habituel (soit à un risque d'erreur de 5%, pour une valeur du test de Student supérieure en valeur absolue à 2). Les variables climatiques ont néanmoins été systématiquement conservées lorsque significatives à un seuil inférieur (à un risque d'erreur voisin de 30%, pour une valeur du test de Student supérieure en valeur absolue à 1).

Sur la période 1975-1993, l'hypothèse $\lambda=0$ est acceptée pour les 6 équations, et l'hypothèse $\lambda=1$ est rejetée 5 fois sur 6.

En ce qui concerne la stabilité du modèle avec forme fonctionnelle optimale, par extension à la période 1975-1998, tous les paramètres sont stables, à l'exception du paramètre de Box-

Cox sur la stabilité duquel l'on ne peut pas se prononcer. L'hypothèse $\lambda=0$ est toujours acceptée pour les 6 équations, et l'hypothèse $\lambda=1$ est maintenant rejetée pour les 6 équations.

Pour un même indicateur, les deux modèles à forme fonctionnelle optimale et à forme contrainte $\lambda=0$ sont statistiquement équivalents, et l'on est naturellement conduit à considérer leurs implications. Nous avons vérifié que ces implications n'étaient pas contradictoires.

Nous nous intéressons ici plus particulièrement à l'interprétation des paramètres estimés avec le modèle à forme fonctionnelle optimale, tant pour les paramètres λ et β_1 de la variable de trafic que pour les paramètres β_j , pour $j=1$ à 4, des variables climatiques.

Les figures 10.1 et 10.2 donnent les différentes formes possibles de la transformation de Box-Cox, et de la fonction d'élasticité de l'endogène à la variable principale, pour différentes valeurs du paramètre λ (-2, 0, 0,5 et 2). Les modèles avec forme fonctionnelle optimale fournissent dans tous les cas, sur la période initiale et sur la période étendue, une estimation négative du paramètre de Box-Cox (celle-ci s'étale entre -0,04 et -1,98), interprétable de la manière suivante : à météorologie donnée, la courbe reliant le nombre d'accidents ou le nombre de tués au trafic est une fonction croissante du trafic, et possède une asymptote. Un effet de saturation des nombres d'accidents et de tués relativement au trafic apparaît au voisinage de l'asymptote, pour des valeurs petites de λ et des valeurs élevées du trafic.

Les élasticités des nombres d'accidents et de tués au trafic sont données par βX^λ ; elles sont calculées pour un mois donné, ou en moyenne annuelle pour une année donnée, comme par exemple pour les années 1975/1984/1993 et 1975/1986/1998 suivant que la période de calage est 1975-1993 ou 1975-1998.

Les élasticités des nombres d'accidents et de tués au trafic sont positives et inférieures à 1, et on retrouve là les ordres de grandeur habituellement proposés, sur données françaises agrégées⁶⁰ Dans tous les cas, elles sont décroissantes vers 0, lorsque le trafic augmente. A

⁶⁰ Comme pour les autres paramètres, les coefficients concernés doivent être examinés avec précaution car ils intègrent une évolution des mécanismes de base et une mesure d'effet direct à court terme. C'est ainsi que les élasticités apparentes du nombre d'accidents au trafic sont plus élevées sur autoroutes que sur routes nationales. Un report de trafic autoroutier vers la route aurait ainsi un impact favorable sur le nombre des accidents ; on peut

titre d'exemple d'amplitude atteinte sur la période, l'élasticité des nombres de tués au trafic, sur les routes nationales, diminue en moyenne annuelle, de 1 en 1975 à 0,55 en 1998, mais à l'intérieur d'une même année elle varie aussi : en 1998 par exemple, elle est de 0,44 en été lorsque le trafic prend les valeurs les plus élevées, et de 0,72 en hiver.

Les paramètres des variables climatiques s'interprètent comme des effets directs sur les indicateurs de risque et de gravité, à niveau de trafic donné. Les ordres de grandeur donnés résultent des unités de mesure des variables climatiques définies en annexe de ce chapitre.

La pluie et l'occurrence de gel, qui réduisent la mobilité, ont néanmoins un effet direct positif (c'est-à-dire un effet d'augmentation) sur les indicateurs de risque d'accident et de gravité. La pluie a un effet statistiquement significatif, au seuil habituel de 95% ($t\text{-ratio} > 2$) pour 4 indicateurs sur les 6 : 100 mm additionnels de hauteur de pluie dans le mois induisent une augmentation de 0,4% à 1% des indicateurs mensuels. L'occurrence de gel a un effet significatif ou moyennement significatif au seuil de 70% ($t\text{-ratio} > 1$) pour 5 indicateurs sur les 6 : 1 jour de gel additionnel dans le mois induit une augmentation comprise entre 1,1% et 1,6% des indicateurs sur les autoroutes, et de l'ordre de 0,1% seulement sur les routes nationales.

Enfin, la température moyenne⁶¹ a un effet très significatif, et particulièrement élevé sur les autoroutes concédées. Les paramètres relatifs à la température maximale sont tous significatifs au seuil habituel sur ce réseau : 1 degré additionnel de température dans le mois induit une augmentation de l'ordre de 3% des nombres d'accidents, et de l'ordre de 4% des nombres de tués, que ce soit en été (avril à septembre) ou en hiver (octobre à mars). Sur l'ensemble des autoroutes, ces augmentations sont moins importantes, et sont de l'ordre de 2% seulement pour les nombres d'accidents. Sur les routes nationales, l'effet de la température est sensiblement atténué, et plus significatif en hiver qu'en été, l'impact correspondant étant compris entre 0,5% et 1,3% en hiver selon l'indicateur considéré.

penser qu'une valeur d'élasticité plus faible sur les routes nationales intègre aussi l'effet des aménagements progressifs qui ont accompagné l'augmentation du trafic.

⁶¹ Il s'agit là d'un effet direct apparent de la température moyenne sur les nombres d'accidents et de tués, puisque les variables de température moyenne et de trafic sont corrélées. Cette remarque vaut en particulier pour la variable de température.

Figure 10.1 : Formes de la fonction de Box-Cox, pour différentes valeurs de $\lambda(\beta>0)$

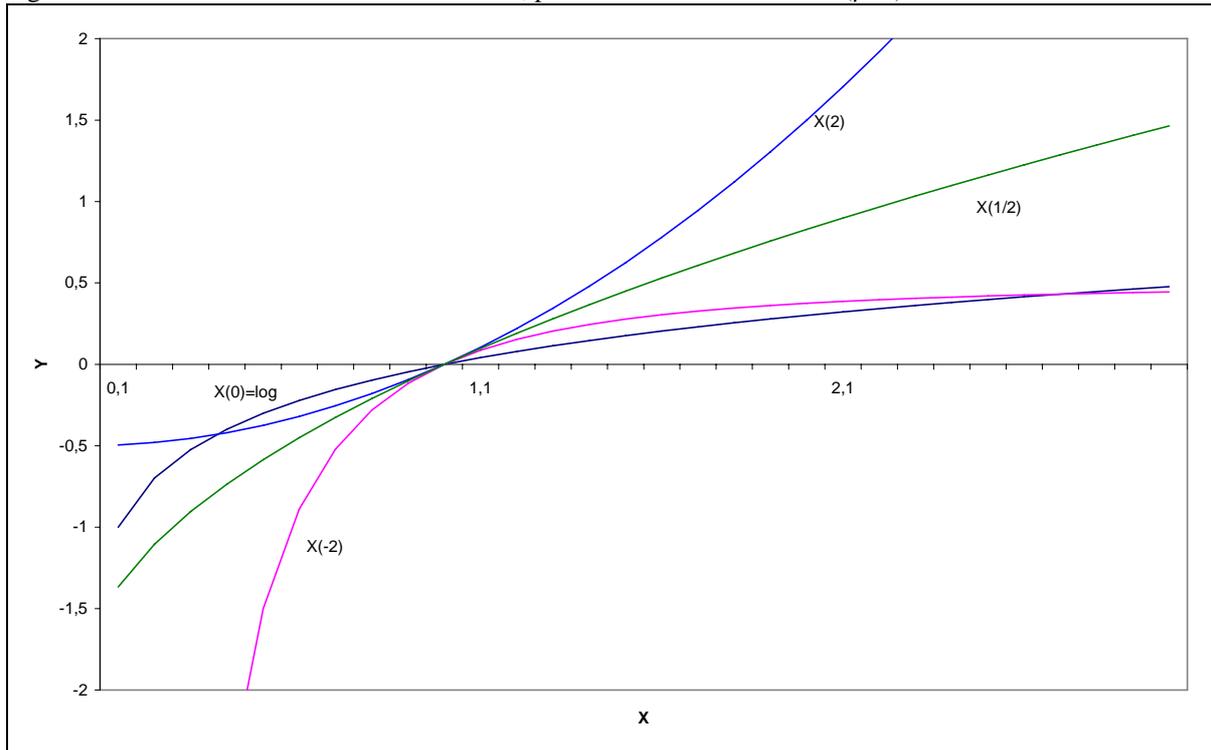


Figure 10.2: Formes de l'élasticité, de Y par rapport à X, pour différentes valeurs de λ ($\beta>0$)

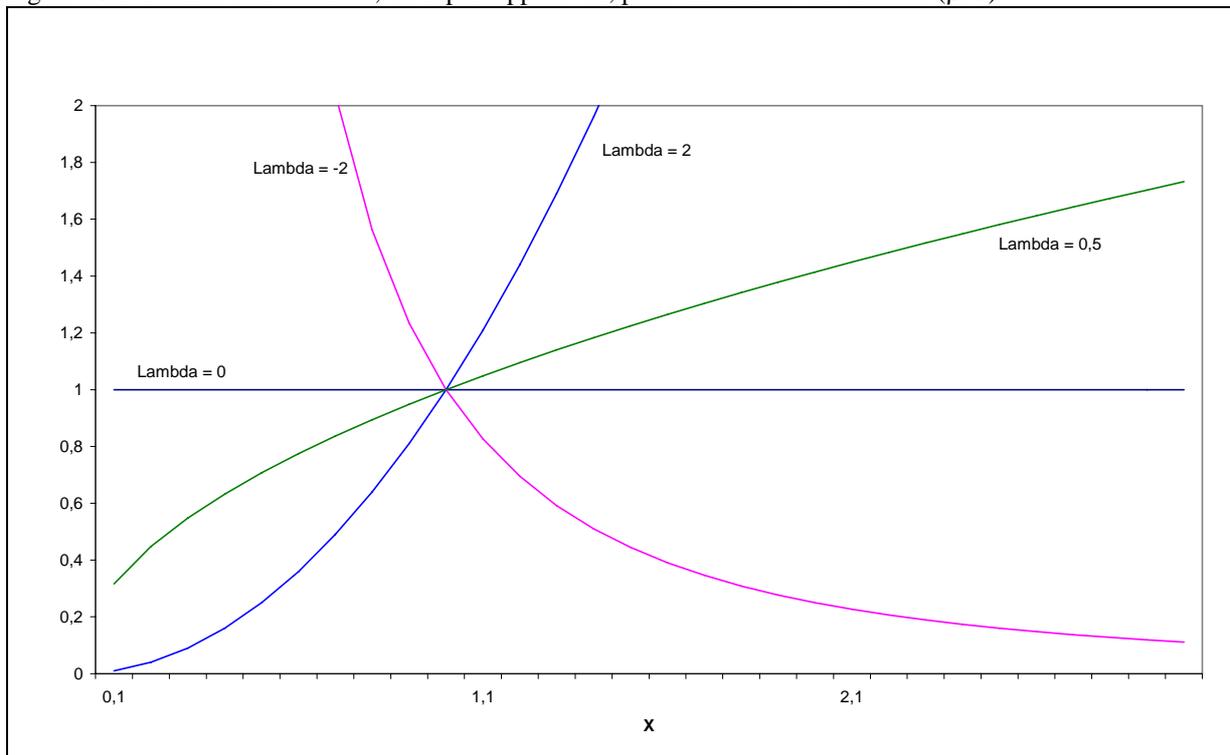


Tableau 10.1 : Modèles avec forme fonctionnelle optimale pour l'exogène principale, sur 1975-1993

	NOMBRES D'ACCIDENTS CORPORELS			NOMBRES DE TUES		
	Autoroutes	Aut. concédées	Routes nationales	Autoroutes	Aut. concédées	Routes nationales
Dynamique						
A(1)	-0,41 ***	-0,27 ***	-0,47 ***			-0,34 ***
A(2)			-0,21 ***			-0,33 ***
A(12)	-0,37 ***	-0,38 ***	-0,55 ***	-0,20 ***	-0,19 ***	-0,40 ***
A(14)	0,24 ***	0,24 ***	0,24 ***			0,10 **
Exogènes (*)						
Trafic autoroutes	1,181309 ***			1,604658 ***		
	-0,28 0,51/0,46/0,36			-0,36 0,55/0,44/0,05		
Trafic autoroutes concédées		0,465353 ***			1,347338 ***	
		-0,04 0,43 / 0,42 / 0,41			-0,36 0,60 / 0,46 / 0,36	
Trafic routes nationales			1,812546 ***			164,155522 ***
			-0,28 0,61 / 0,59 / 0,55			-1,33 0,95 / 0,82 / 0,64
Hauteur de pluie	0,000076 ***	0,000083 ***	0,000065 ***	0,000003 *	0,000001 *	0,000080 ***
Jours de gel	0,008152 ***	0,014245 ***	0,001493 *	0,010105 **	0,014493 **	0,006577 **
Température été	0,001517 ***	0,003130 ***	0,000213 *	0,003344 ***	0,004139 ***	0,000038 *
Température hiver	0,001858 ***	0,003006 ***	0,000602 **	0,003944 ***	0,004475 ***	0,001063 ***
Qualité du modèle						
Log vraisemblance	221,775	170,166	286,009	2,801	-65,111	170,851
SSE	1,882	2,935	1,053	12,997	23,544	2,917
R ²	52,1%	68,6%	33,1%	46,0%	45,5%	27,0%
Tests sur les résidus						
Echantillonnage	accepté	rejeté	accepté	rejeté	accepté	rejeté
P-value	0,30	0,01	0,56	0,02	0,12	0,04
Normalité	accepté	accepté	rejeté	accepté	accepté	accepté
P-value	0,39	0,54	0,004	0,88	0,92	0,34
Bruit blanc	accepté	accepté	accepté	accepté	accepté	accepté
P-value	0,06	0,61	0,14	0,07	0,64	0,10
Test de comparaison						
H0:lambda=0 contre H1:lambda<>1	LV0=221,371 H0 acceptée	LV0=170,118 H0 acceptée	LV0=285,942 H0 acceptée	LV0=2,562 H0 acceptée	LV0=-65,416 H0 acceptée	LV0=169,615 H0 acceptée
H0:lambda=1 contre H1:lambda<>	LV0=214,893 H0 rejetée	LV0=161,642 H0 rejetée	LV0=284,857 H0 acceptée	LV0=-0,35616 H0 rejetée	LV0=-69,242 H0 rejetée	LV0=167,475 H0 rejetée

Significativité : *(t-ratio<1), ** (1<t-ratio<2), *** (t-ratio>2)

Le tableau contient :

- pour la dynamique et les exogènes secondaires : le paramètre et sa significativité,

- pour l'exogène principale: le paramètre, sa significativité, la valeur de λ et les élasticités, en moyenne annuelle en 1975, 1984 et 1993 en gras.

Tableau 10.2 : Modèles avec forme fonctionnelle optimale pour l'exogène principale, sur 1975-1998

	NOMBRES D'ACCIDENTS CORPORELS			NOMBRES DE TUES		
	Autoroutes	Aut. concédées	Routes nationales	Autoroutes	Aut. concédées	Routes nationales
Dynamique						
A(1)	-0,41 ***	-0,26 ***	-0,49 ***			-0,34 ***
A(2)			-0,18 ***			-0,29 ***
A(12)	-0,43 ***	-0,49 ***	-0,54 ***	-0,25 ***	-0,21 ***	-0,40 ***
A(14)	0,21 ***	0,17 ***	0,21 ***	-0,14 ***		0,03 *
Exogènes (*)						
Trafic autoroutes	1,450242 ***			273,149373 ***		
	-0,36 0,50/0,38/0,29			-1,98 0,79/0,18/0,05		
Trafic autoroutes concédées		0,841966 ***			10,277121 ***	
		-0,28 0,45/0,35/0,28			-1,17 0,71/0,25/0,11	
Trafic routes nationales			4,737984 ***			120,725051 ***
			-0,53 0,62/0,57/0,49			-1,25 0,96/0,76/0,55
Hauteur de pluie	0,000103 ***	0,000105 ***	0,000069 ***	0,000041 *	0,000048 *	0,000081 ***
Jours de gel	0,010644 ***	0,014605 ***	0,001354 *	0,013871 **	0,015804 **	0,007500 ***
Température été	0,001639 ***	0,003032 ***	0,000120 *	0,003604 ***	0,004382 ***	0,000534 **
Température hiver	0,001911 ***	0,002767 ***	0,000456 **	0,003845 ***	0,004506 ***	0,001325 ***
Qualité du modèle						
Log vraisemblance	278,806	215,670	358,918	-7,7464	-82,679	202,032
SSE	2,402	3,711	1,295	17,731	29,678	3,940
R ²	46,0%	53,5%	32,0%	28,6%	39,3%	24,4%
Tests sur les résidus						
Echantillonnage	accepté	accepté	accepté	accepté	accepté	accepté
P-value	0,56	0,05	0,44	0,89	0,66	0,10
Normalité	accepté	accepté	rejeté	accepté	accepté	rejeté
P-value	0,44	0,98	0,0004	0,54	0,95	0,03
Bruit blanc	accepté	accepté	rejeté	accepté	accepté	rejeté
P-value	0,05	0,29	0,01	0,15	0,46	0,003
Test de comparaison						
H0:lambda=0 contre H1:lambda<>1	LV0=277,828 H0 acceptée	LV0=210,009 H0 acceptée	LV0=358,536 H0 acceptée	LV0=-10,1102 H0 acceptée	LV0=-83,582 H0 acceptée	LV0=200,447 H0 acceptée
H0:lambda=1 contre H1:lambda<>	LV0=272,507 H0 rejetée	LV0=202,566 H0 rejetée	LV0=356,324 H0 rejetée	LV0=-12,872 H0 rejetée	LV0=-87,870 H0 rejetée	LV0=197,591 H0 rejetée

Significativité : *(t-ratio<1), ***(1<t-ratio<2), ***(t-ratio>2)

Le tableau contient :

- pour la dynamique et les exogènes secondaires : le paramètre et sa significativité,

- pour l'exogène principale: le paramètre, sa significativité, la valeur de λ et les élasticités, en moyenne annuelle en 1975, 1986 et 1998 en gras

10.5. Le modèle étendu

Comme déjà indiqué, la modélisation qui vient d'être exposée a été étendue à deux autres réseaux : réseau secondaire et milieu urbain, et à la France entière. Pour ces trois réseaux, les variables de trafic manquent et ont été remplacées par des variables représentatives de l'offre et de la demande de transport. Lors de cet exercice d'extension, les variables météorologiques, déjà mesurées en moyennes mensuelles, ont été assorties de variables d'atypicité météorologique témoins de valeurs journalières extrêmes, et de variables calendaires. Nous avons généralisé la spécification économétrique contrainte ressortant de l'exercice précédent, sans relancer la recherche de forme fonctionnelle qui vient d'être exposée.

10.5.1. Objectif

Rappelons que l'objectif consiste à évaluer les effets des facteurs conjoncturels de nature climatique et calendaire sur les nombres d'accidents et de décès enregistrés sur la route, agrégés sur la France entière et désagrégés par grande catégorie de réseau (routes nationales, autoroutes, réseau secondaire, milieu urbain).

Sur les réseaux pour lesquels les variables de trafic manquent ; nous avons cherché à quantifier un effet conjoncturel global, de nature climatique ou calendaire, sur chaque indicateur modélisé. Sur le réseau routier national (routes nationales et autoroutes) sur lequel le trafic est mesuré avec une périodicité fine, nous avons cherché à décomposer l'effet conjoncturel global en un effet indirect qui se manifeste sur chaque indicateur par le biais du trafic, indicateur d'exposition au risque et principal facteur de risque, et un effet direct à niveau de trafic inchangé.

10.5.2. Méthode

10.5.2.1. Structure du modèle

Le modèle est établi sur une base mensuelle, et sur la période 1975-1999, et s'applique aux *nombres mensuels d'accidents corporels et de tués, agrégés sur la France entière et désagrégés par catégorie de réseau* (routes nationales, autoroutes, réseau secondaire, milieu urbain).

Les facteurs explicatifs pris en compte sont d'une part l'exposition au risque, mesurée par *le volume de trafic* sur le réseau routier national - ou par ses déterminants, qui tiennent lieu de proxy du trafic, sur les autres réseaux -, et d'autre part *le facteur climatique et calendaire*.

Les variables météorologiques retenues pour modéliser l'influence du climat sur les indicateurs de sécurité routière mesurent *la température maximale du jour, l'occurrence de gel (température négative) et la hauteur de pluie*, moyennées sur la France entière et sur le mois. Elles sont dans un deuxième temps assorties de variables « d'atypicité météorologique », témoins de valeurs journalières extrêmes. Des variables calendaires décomptent les jours de fin de semaine (ce découpage lundi-jeudi/fin de semaine permet de modéliser un effet de *calendrier hebdomadaire*), ainsi que trois classes de jour témoins d'une modification du comportement du conducteur (ce découpage permet de modéliser un effet de *calendrier exceptionnel*).

10.5.2.2. Spécification économétrique

Nous avons généralisé la spécification économétrique contrainte ressortant de l'exercice précédent, en raison de sa parcimonie, de la manière suivante : les variables exogènes mesurant l'exposition au risque ou ses déterminants ont subi une transformation logarithmique (ce qui reviendrait, pour chacune d'elles, à contraindre à 0 le paramètre λ d'une transformation de Box-Cox) et interviennent donc de manière multiplicative, et les variables exogènes mesurant les conditions climatiques n'ont pas été transformées et interviennent de manière semi additive). Il ressort de la revue bibliographique des modèles agrégés de risque routier développés au niveau national en Europe, et présentée dans le chapitre 8, que cette spécification est très largement utilisée.

Le modèle retenu pour chaque indicateur est le suivant :

$$\Phi(B) \left[\log Y_t - \sum_{i=1}^I \beta_i \log X_{i,t} - \sum_{j=1}^J \beta_j X_{j,t} - \mu \right] = \Theta(B) u_t$$

avec : Y la variable endogène : nombre d'accidents corporels et nombre de tués,

X_i $i=1, \dots, I$ les I variables exogènes principales, mesurant l'exposition au risque ou ses déterminants⁶²,

X_j $j=1, \dots, J$ les J variables explicatives secondaires mesurant les conditions climatiques,

⁶² Sur le réseau routier national (routes nationales et autoroutes) sur lequel le trafic est mesuré, le modèle a aussi été établi pour les indicateurs de trafic : on peut ainsi sur ces deux réseaux déterminer les effets indirects des variables conjoncturelles sur les nombres d'accidents et de tués par le biais du trafic, en complément des effets directs donnés par les modèles explicatifs des nombres d'accidents et de tués.

$\Phi(B)$ et $\Theta(B)$ deux polynômes en B , représentant la dynamique de Y , B l'opérateur retard,

et u_t un bruit blanc non corrélé avec le passé de Y , des X_i $i=1,\dots,I$ et des X_j $j=1,\dots,J$.

Les variables, endogène et exogènes, sont obtenues par application d'un filtre $(I-B)^{12}$ aux données initialement observées. Cette transformation garantit, dans tous les cas, la stationnarité de la variable endogène transformée, et corrigée des effets exogènes.

10.5.2.3. Estimation

L'estimation des paramètres est réalisée par le maximum de vraisemblance, en supposant que la variable endogène transformée, conditionnellement aux variables exogènes, est gaussienne. Cette hypothèse de normalité se vérifie sur le résidu du modèle, mais elle n'est pas fondamentale : si l'hypothèse de normalité n'est pas validée, le calcul de la vraisemblance peut être remis en cause, mais les estimateurs peuvent néanmoins avoir de bonnes propriétés de convergence asymptotique.

10.5.3. Résultats

Sont successivement donnés les principaux éléments de validation et d'évaluation empirique du modèle, puis une interprétation des paramètres liés à l'exposition au risque puis aux variables conjoncturelles. Les interprétations des effets de chaque variable exogène s'entendent « toutes choses égales par ailleurs », ce qui suppose que les variables exogènes sont non corrélées. Cette hypothèse de non corrélation peut être retenue entre variables climatiques et calendaires, mais elle est moins valide s'agissant par exemple du trafic et de la température.

10.5.3.1. Validation/évaluation

Les tests de validation les plus importants ont été réalisés, et sont satisfaisants : les principaux paramètres, des polynômes $\Phi(B)$ et $\Theta(B)$, et des variables exogènes $X_{i,t}$, sont significatifs au seuil habituel de 95%.

Les paramètres des variables exogènes secondaires $X_{j,t}$ ne sont pas toujours significatifs au seuil habituel, mais les variables secondaires ont néanmoins été systématiquement conservées pour des raisons de commodité.

L'hypothèse principale relative au résidu u_t , de non corrélation, est toujours validée.

L'évaluation du modèle est réalisée à l'aide de critères de performance empirique.

Sont donnés : le critère d'information d'Akaike, la part de variance expliquée (sur les endogènes corrigées des effets exogènes, et préalablement filtrées par le filtre (I-B¹²)), et enfin une mesure de l'erreur moyenne d'ajustement faite sur la période : le MAPRE⁶³.

La performance varie fortement selon l'indicateur et selon le réseau. L'erreur moyenne d'ajustement sur la période étudiée est de l'ordre de 3% pour le nombre d'accidents France entière, et de 5% pour le nombre de tués France entière. Mais pour les indicateurs décomposés par réseau, elle augmente très sensiblement, avoisinant par exemple 7% et 20% sur les autoroutes, et même 8% et 25% sur les autoroutes concédées. Ces différences de performance entre les modèles s'expliquent d'abord par les différences de taille des variables endogènes (sur la période de calage, la moyenne du nombre mensuel de tués est de .. seulement sur les autoroutes), et par l'absence de facteurs spécifiques, non mesurés ou même non mesurables, du risque d'accident et de sa gravité, par catégorie de réseau.

10.5.3.2. Exposition au risque

Sur le réseau secondaire, en agglomération et sur la France entière, réseaux pour lesquels l'on ne dispose pas d'une mesure mensuelle du trafic, la relation au volume de trafic ne peut pas être quantifiée directement. L'on a ici recours à des variables représentatives de l'offre et de la demande agrégée de transport, qui sont les déterminants du trafic : le prix des carburants et la consistance du réseau pour l'offre, et la consommation des ménages pour la demande de transport. En raison de la corrélation de ces variables – et en particulier de la consistance du réseau routier et de la consommation des ménages -, il n'est pas possible de séparer leurs effets respectifs sur les indicateurs de sécurité routière, ni donc d'interpréter leurs paramètres dans le modèle comme des élasticités, mais seulement comme des élasticités apparentes.

Sur routes nationales et sur autoroutes, la relation du nombre d'accidents et du nombre de tués au volume de trafic est mesurée par une élasticité constante sur la période.

Les élasticités apparentes des nombres d'accidents et de tués au trafic sont de l'ordre de 0,5 sur les routes nationales ; sur les autoroutes, elles seraient avec la spécification de court terme retenue ici de l'ordre de 0,9 pour le nombre d'accidents et de 1,6 à 1,8 pour le nombre de tués. Cependant, en raison de la corrélation entre la température et le trafic, les effets respectifs de ces deux variables sur le nombre d'accidents et de tués sont difficiles à séparer ; les estimations de la relation des nombres d'accidents et de tués au trafic restent à approfondir.

10.5.3.3. Effets climatiques

Rappelons que les ordres de grandeur donnés pour ces effets résultent des unités de mesure des variables climatiques définies en annexe de ce chapitre.

(i) Variables météorologiques moyennes mensuelles (cf. Tableaux 10.3 à 10.4)

Température moyenne :

La température est liée, de manière positive, au nombre total d'accidents.

Ainsi, une augmentation d'un degré de la température mensuelle moyenne se traduit par une augmentation de l'ordre de 1% du nombre total d'accidents (0,8% l'été, et 1,2% l'hiver). De plus, l'hiver, l'occurrence de gel (température négative) intervient également, comme nous le verrons plus loin.

Cet effet de la température sur le nombre d'accidents s'observe sur tous les réseaux, mais de manière beaucoup plus importante sur les autoroutes.

Ainsi, une augmentation d'un degré de la température en été se traduit par une augmentation du nombre d'accidents sur autoroutes de l'ordre de 2% (et de 2,4% sur les autoroutes concédées). Il s'agirait sur ce réseau d'un effet essentiellement direct, c'est-à-dire d'une augmentation du nombre d'accidents à niveau de trafic inchangé, (soit +1,7% et +2,2% sur les autoroutes concédées), qui met en évidence une prise de risque de la part du conducteur. L'effet indirect, du à l'augmentation du trafic (soit +0,3% et +0,2% respectivement), est quant à lui tout à fait limité au regard de l'effet direct.

Sur les routes nationales, l'effet de la température est très atténué, et serait pour une grande part indirect. Mais sur le réseau secondaire et en agglomérations, il est du même ordre que l'effet moyen France entière (+0,8% l'été et +1,2% l'hiver sur le réseau secondaire, et +0,6% et +1,2% respectivement en agglomérations).

La température est également liée, de manière positive, au nombre total de décès.

Ainsi, une augmentation d'un degré de la température mensuelle moyenne se traduit par une augmentation de l'ordre de 1% du nombre total de tués en été, et de 2% en hiver.

L'analyse par réseaux montre une relation significative, entre la température moyenne et le nombre de tués sur le réseau secondaire et en agglomérations, qui est du même ordre que

⁶³ Mean Average Percentage Relative Error, ou erreur moyenne relative d'ajustement

l'effet moyen France entière. Comme dans le cas des accidents, la relation au nombre de tués serait très marquée sur les autoroutes (plus de 4% d'effet global sur les autoroutes concédées, qui serait essentiellement direct) alors qu'elle est à nouveau du même ordre que l'effet moyen sur les routes nationales, l'effet étant sur ce réseau essentiellement direct – contrairement à ce que l'on observe dans le cas des accidents, pour lesquels l'effet indirect lié à la diminution du trafic est important⁶⁴.

En raison de la corrélation entre le volume de trafic et la température moyenne, la séparation qui vient d' être donnée, sur routes nationales et sur autoroutes, entre effet météorologique direct et indirect par le biais du trafic reste à approfondir.

Occurrence de gel moyenne :

L'occurrence de gel est liée, de manière négative, au nombre total d'accidents.

Un jour de gel dans le mois se traduit par une diminution de 0,3% du nombre mensuel d'accidents.

La diminution est du même ordre de grandeur sur le réseau secondaire et en agglomérations.

Sur les autoroutes, cette relation est inversée : on enregistre une augmentation de 0,6% et de 1,1% du nombre d'accidents sur autoroutes et sur autoroutes concédées pour un jour de gel dans le mois, et cet effet est essentiellement direct (+0,7% et +1,2% respectivement), traduisant une augmentation du risque alors que l'effet indirect, lié à la diminution du trafic, est négligeable.

Sur les routes nationales, seul l'effet indirect, négatif, apparaît moyennement significatif.

La relation de l'occurrence de gel au nombre de tués ne s'observe, de manière moyennement significative, qu'en agglomération (-0,5% pour un jour de gel dans le mois). La relation est à nouveau inversée sur les autoroutes et sur les routes nationales, où l'effet est essentiellement direct (+1,1% sur les autoroutes, +1,8% sur les autoroutes concédées, et plus de 3% sur les routes nationales - cette dernière estimation, toutefois, est peu significative), traduisant à nouveau une augmentation du risque.

Hauteur de pluie moyenne :

La hauteur de pluie est liée, de manière positive, au nombre total d'accidents et de tués.

64

Ainsi, 100 mm de hauteur de pluie dans le mois se traduisent par une augmentation de 0,3% du nombre d'accidents et de 0,2% du nombre de tués du mois.

Sur les routes nationales et sur les autoroutes, on enregistre un effet direct positif, très significatif, sur les nombres d'accidents et de tués (une augmentation de 10 mm de hauteur de pluie se traduit par une augmentation de 0,1 à 0,2% du nombre d'accidents et de tués selon l'indicateur), qui met en évidence une augmentation du risque, et qui dépasse très largement l'effet indirect négatif lié à la diminution du trafic.

(ii) Variables météorologiques moyennes mensuelles et variables d'atypicité météorologique (cf. Tableaux 10.5 et 10.6)

Cinq variables supplémentaires, qualifiées « d'atypicité météorologique », sont maintenant prises en compte. Il s'agit d'indicatrices du nombre de jours du mois pour lesquels la variable météorologique journalière a pris des valeurs extrêmes, en s'écartant significativement de la normale saisonnière journalière.

Les effets climatiques mensuels se décomposent maintenant entre un effet moyen – moins significatif que l'effet de moyenne mensuelle décrit auparavant - et un effet d'atypicité. Seuls les effets additionnels d'atypicité, lorsqu'ils sont significatifs, sont commentés ci-après.

Atypicité en température :

De manière générale, lorsque l'on prend également en compte les variations extrêmes de température (atypicité supérieure en température d'été, atypicité inférieure et supérieure en température d'hiver), la relation à la température moyenne apparaît moins significative l'hiver que l'été. En hiver, la température extrême – grand froid -, intervient également, et de manière plus significative que la température moyenne.

L'atypicité inférieure en température d'hiver est liée, de manière négative, au nombre total d'accidents et de tués.

Un jour d'atypicité inférieure en température d'hiver se traduit par une diminution de 0,9% du nombre d'accidents et de 1,2% du nombre de tués.

La diminution du nombre d'accidents est du même ordre en agglomération, et légèrement moindre sur le réseau secondaire ; sur les routes nationales et sur les autoroutes, seul un effet indirect négatif, lié à la diminution du trafic (-0,3% par jour d'atypicité), est significatif.

La diminution du nombre de tués s'observe sur tous les réseaux, mais de manière moins significative ; sur les routes nationales et sur les autoroutes, à l'effet négatif lié à la diminution

du trafic (-0,2% et -0,6% respectivement) vient s'ajouter un effet direct négatif (-0,6% et -2,8%), mais seul l'effet indirect est significatif.

L'atypicité supérieure en température d'hiver n'est pas liée au nombre total d'accidents et de tués. Mais sur le réseau secondaire cette liaison est importante (+0,4% et +1,1% respectivement par jour d'atypicité supérieure), On peut aussi mentionner, bien qu'il soit moyennement significatif, un effet direct positif sur le nombre de tués sur les autoroutes, qui vient s'additionner à un effet indirect bien plus limité (+2,9% et +0,3% respectivement).

L'atypicité supérieure en température d'été n'est pas non plus liée au nombre total d'accidents et de tués.

Cependant, elle est liée, positivement, au nombre de tués en agglomération (+4,4% du nombre mensuel de tués par jour d'atypicité supérieure en température d'été).

Atypicité en occurrence de gel :

L'atypicité supérieure en gel (gel fort) est liée, de manière négative mais moyennement significative, au nombre total de tués. Sur les routes et les autoroutes, à l'effet indirect négatif lié à la diminution du trafic s'ajoute un effet direct positif sur le nombre de tués sur autoroutes (+1,3% de tués pour un jour de gel fort dans le mois sur les autoroutes, et +3,1% sur les autoroutes concédées).

Atypicité en pluie :

L'atypicité supérieure en pluie se traduit par un effet direct négatif sur les nombres d'accidents et de tués sur autoroutes (soit -0,6% d'accidents et -4% de tués sur autoroutes) qui traduit une diminution de la prise de risque - l'effet indirect de diminution du trafic n'étant pas significatif.

10.5.3.4. Effets calendaires

Notons, pour les deux indicateurs agrégés l'impact d'un jour de fin de semaine (+0,3% et +1% du nombre mensuel d'accidents et du nombre mensuel de tués France entière pour un jour supplémentaire de fin de semaine dans le mois), l'impact des veilles de fériés et des grands déplacements liés aux congés (+0,4% et +0,7% respectivement) et celui des jours fériés (-0,3% et +0,3% respectivement).

Il est plus difficile de mettre en évidence des effets calendaires sur les nombres d'accidents et de tués décomposés par catégorie de réseau, car la classification des jours de l'année témoins d'un effet calendaire est à mener différemment selon le réseau.

Cependant, sur les routes nationales et les autoroutes, l'effet calendaire indirect sur les nombres d'accidents et de tués par le biais du trafic a pu être estimé. Ces effets indirects sont particulièrement élevés (entre +0,1% et +1,6% du nombre mensuel d'accidents pour un jour supplémentaire de trafic élevé ou très élevé, selon le réseau, et entre +0,1% et +2,6% du nombre mensuel de tués respectivement), les corrélations minimales sont obtenues sur les routes nationales.

10.5.3.5. Analyse de l'effet météorologique en 1999 et en 2000

Nous résumons ici les sorties du modèle explicatif de l'évolution des nombres mensuels d'accidents et de tués pour la France entière, pour la partie relative au facteur climatique et pour les années 1999 et 2000, qui ont été globalement défavorables à la sécurité routière.

Les mois de 1999 et 2000 qui ont été témoins d'un effet météorologique important sur la sécurité routière, mesurée par les nombres d'accidents et de décès, sont détaillés ci-après ; les écarts des variables météorologiques quantitatives utilisées dans le modèle à leur normale saisonnière, sont également donnés, ainsi que l'effet météorologique qui en résulte, mesuré en nombre d'accidents et de décès.

L'ordre de grandeur de l'effet météorologique est par exemple de 1%, en moyenne sur l'année 2000, de la valeur mensuelle de l'indicateur modélisé, qu'il s'agisse du nombre d'accidents ou du nombre de tués. Ce résultat est à mettre en perspective avec l'ordre de grandeur de l'effet mensuel qui dépasse les 10 % certains mois, et avec celui de l'effet journalier qui atteint des ordres de grandeur sensiblement plus élevés, comme nous l'avons vu en 9.3.1 avec les sorties du modèle Giboulée.

Année 1999 : La météorologie a été globalement défavorable sur l'année 1999, avec un effet total de l'ordre de 840 accidents et de 37 tués (à ramener aux 124 524 accidents et 8029 tués enregistrés en 1999).

Au total en 1999, 5 mois ont été témoins d'un effet météorologique important. D'abord, janvier et septembre, et dans une moindre mesure décembre, ont connu une température particulièrement douce, à l'opposé février et surtout novembre, avec 2 jours de température particulièrement faible, ont connu une température plus faible. Ensuite, septembre et décembre ont été pluvieux. Enfin, l'occurrence de gel a été forte en février.

Résumé par mois :

Janvier : température plus élevée (+ 2 °C)⁶⁵,

Février : température plus faible (- 1 °C) et 5 jours de gel fort⁶⁶,

Septembre : plus de pluie (+ 450 mm/normale)⁶⁷ et température plus élevée (+2°C),

Novembre : température plus faible (-1,4°C) avec 2 jours de température très faible⁶⁸,

Décembre : plus de pluie (+ 600 mm/normale), température plus élevée (+0,5°C).

Tous effets confondus, pour les accidents, les mois favorables ont été février, novembre et dans une moindre mesure juin (-110, -177 et -61 accidents respectivement), alors que tous les autres ont été défavorables et en particulier janvier, septembre et décembre (+166, +250 et +277 accidents).

Ensuite pour les tués, les mois favorables sont février, juin et novembre (-16, -3 et -18 tués respectivement), alors que tous les autres étaient défavorables et notamment janvier, septembre et décembre (+13, +14, +13 tués).

Année 2000 : La météorologie a été globalement défavorable sur l'année 2000, avec un effet total de l'ordre de 1100 accidents et de 66 tués (à ramener aux 121 223 accidents et 7643 tués enregistrés en 2000).

Au total en 2000, 4 mois ont été témoins d'un effet météorologique important.

D'abord, les mois de décembre et février, et dans une moindre mesure novembre, ont connu une température particulièrement douce. Ensuite, novembre a été pluvieux et janvier plus sec.

Enfin, l'occurrence de gel a été faible en février, novembre et décembre, et forte en janvier.

Résumé par mois :

Janvier : moins de pluie (-500 mm/normale) et 4 jours de gel fort,

Février : température plus élevée (+2,3 °C) et moins de jours de gel (-4 j/normale)⁶⁹,

Novembre : plus de pluie (+700 mm/normale) et température plus élevée (+1 °C) et moins de jours de gel (-5j/normale),

Décembre : température plus élevée (+3 °C) et moins de jours de gel (-7 j/normale)

Tous effets confondus, pour le nombre d'accidents, janvier est favorable (-260 accidents) alors que la plupart des autres mois sont défavorables, et notamment février, novembre, décembre (+218, +376 et + 278 accidents respectivement). Ensuite pour le nombre de tués, janvier est favorable (-18 tués), alors que la plupart des autres mois sont défavorables, et notamment février, novembre et décembre (+18, +19, +22 tués respectivement).

⁶⁵ + 1 ° C/normale saisonnière du mois occasionne 72 accidents et 8 tués en hiver,
+ 1 ° C/normale saisonnière du mois occasionne 28 accidents et 2 tués en été.

⁶⁶ +1 jour atypique de gel dans le mois fait gagner 6 accidents et 1 tué.

⁶⁷ +100 mm/normale saisonnière du mois occasionne 30 accidents et 2 tués.

⁶⁸ +1 jour atypique de température faible dans le mois fait gagner 16 accidents et 2 tués.

⁶⁹ +1 jour de gel dans le mois fait gagner 1 accident et occasionne 0,52 tué.

10.6. Conclusion

Dans un premier temps, nous nous sommes limitée à considérer deux niveaux de risque et de gravité, mesurés par les nombres d'accidents et de tués, et nous avons cherché à valider sur le réseau national (routes et autoroutes), une forme fonctionnelle comportant un paramètre non linéaire. Deux facteurs de risque sont pris en compte : l'exposition au risque mesurée par une variable principale de trafic enregistré sur le réseau, et le facteur climatique mesuré par des variables secondaires relatives à la température, à l'occurrence de gel et à la pluie.

Un modèle avec transformation logarithmique de l'endogène et transformation de Box-Cox de paramètre λ sur l'exogène principale a été établi sur la période 1975-1993. Des tests de comparaison de la spécification initiale à deux cas particuliers ($\lambda=0$ et $\lambda=1$) ont été réalisés, et les réponses à ces tests sont les mêmes sur la période 1975-1993 et sur la période 1975-1998. Il ressort qu'il n'y a pas de différence significative entre le modèle avec transformation de Box-Cox sur l'exogène principale et le modèle avec transformation logarithmique sur l'exogène principale ($\lambda=0$), de sorte que l'on peut pour des raisons de parcimonie retenir cette seconde spécification, très largement utilisée. Ce constat s'applique pour tous les indicateurs modélisés, sur les deux réseaux considérés.

Ce résultat ne permet pas de trancher entre le modèle à forme fonctionnelle optimale et le modèle à forme contrainte $\lambda=0$, et l'on est naturellement amené à considérer leurs implications. Nous avons vérifié que celles-ci n'étaient pas contradictoires.

Dans un deuxième temps, nous avons retenu la spécification contrainte ainsi validée, et nous l'avons étendue à deux réseaux additionnels : réseau secondaire, milieu urbain, et à la France entière. Sur ces réseaux, les variables de trafic manquent et ont été remplacées par des variables représentatives de l'offre et de la demande de transport.

Le modèle étendu a été établi sur la période 1975-1999. L'exposition au risque - quand elle est mesurée - ou ses déterminants -, ainsi que des facteurs de risque conjoncturels, de nature climatique et calendaire, sont pris en compte. Les variables météorologiques mesurent la température, l'occurrence de gel et la hauteur de pluie : des variables d'atypicité climatique - qui codent les valeurs extrêmes prises dans le mois - ont été utilisées en complément des moyennes mensuelles.

Des liaisons significatives apparaissent entre les variables climatiques et les deux indicateurs de risque, pour la France entière et par catégorie de réseau. Pour la France entière, l'effet des conditions météorologiques sur les nombres d'accidents et de tués a été analysé en 1999 et en 2000. Sur les routes nationales et sur les autoroutes, les deux catégories de réseau sur lesquelles le trafic mensuel est mesuré, l'effet global du climat a pu être séparé en deux composantes distinctes : son effet direct sur les nombres d'accidents et de tués, à volume de trafic inchangé, et son effet indirect par le biais du trafic.

En raison de la corrélation entre le volume de trafic et la température moyenne, notamment, la séparation qui a été faite sur routes nationales et sur autoroutes entre effet direct et indirect par le biais du trafic reste à approfondir.

Des variations du niveau de risque ont été mises en évidence dans certaines situations, sous l'influence des variations de ces variables climatiques. Ces résultats agrégés au niveau du mois doivent être analysés plus avant au niveau du jour, et les liens avec les variables relatives au comportement des conducteurs doivent être étudiés, de manière à compléter ces premiers résultats.

Tableau 10.3 : Les paramètres du modèle initial sur données filtrées, sur 1977-1999

France entière, réseau secondaire, agglomérations														
Endogène	AIC	MAPRE_R	R ²	Exogènes										
				LGA	CFM	ICARB	PAAU	PAAC	PARN	TE	TH	HPLUI	NGEL	
Accidents														
ACCFE	-898,7	3,3%	65,6%	-0,07136	0,14396	0,13397					0,00075	0,00121	0,00003	-0,00283
				*	**	**					***	***	***	**
ACCRS	-753,5	4,3%	56,0%		0,04936	0,15490					0,00080	0,00120	0,00001	-0,00231
					*	**					***	***	*	**
ACCAGG	-828,8	3,8%	93,7%		0,23402	0,14259					0,00056	0,00122	0,00002	-0,00362
					**	**					***	***	***	***
Tués														
TUEFE	-672,0	5,1%	58,3%	0,12430	0,52255	0,01522					0,00107	0,00215	0,00002	-0,00071
				*	***	*					***	***	**	*
TUERS	-538,2	6,6%	51,2%		0,39507	0,17751					0,00092	0,00251	0,00002	-0,00027
					**	**					***	***	*	*
TUEAGG	-448,3	8,0%	93,7%		0,80418	-0,04087					0,00118	0,00147	-0,00002	-0,00530
					***	*					***	***	*	**

Autoroutes, autoroutes concédées et routes nationales														
Endogène	AIC	MAPRE_R	R ²	Exogènes										
				LGA	CFM	ICARB	PAAU	PAAC	PARN	TE	TH	HPLUI	NGEL	
Accidents														
ACCA	-522,8	6,9%	54,1%				0,89658				0,00173	0,00072	0,00009	0,00678
							***				***	**	***	***
ACCAC	-419,5	8,2%	44,0%					1,08929			0,00217	0,00137	0,00009	0,01212
								***			***	**	***	***
ACCRN	-735,2	4,7%	51,6%						0,46659	0,00014	0,00003	0,00006	-0,00190	
									***	*	*	***	*	
Tués														
TUEA	39,1	20,5%	38,0%				2,03770				0,00125	0,00186	0,00002	0,01136
							***				*	*	*	**
TUEAC	160,0	26,2%	35,4%					1,77798			0,00032	0,00397	0,00003	0,01831
								***			*	**	*	**
TUERN	-425,9	8,4%	47,9%						0,42106	0,00101	0,00134	0,00007	0,00351	
									***	**	**	***	*	
Parcours														
PAAU	-1061,7	2,3%	38,0%	0,10911	0,17788	-0,11510					0,00028	0,00050	-0,00001	-0,00096
				*	***	***					**	***	**	*
PAAC	-904,4	3,2%	41,2%	0,28189	0,27777	-0,13394					0,00022	0,00038	-0,00002	-0,00130
				**	***	**					**	**	***	*
PARN	-1234,5	1,7%	51,0%		0,13221	-0,06133					0,00031	0,00075	-0,00001	-0,00123
					***	**					***	***	***	**

Significativité des paramètres : ***(t-ratio supérieur à 2), *(t-ratio entre 1 et 2), *(t-ratio inférieur à 1)

Tableau 10.4 : Effets des variables climatiques, sur 1977-1999

		TE	TH	HPLUI	NGEL
Accidents					
ACCFE	global	0,08% ***	0,12% ***	0,00% ***	-0,28% **
ACCRS	global	0,08% ***	0,12% ***	0,00% *	-0,23% **
ACCAGG	global	0,06% ***	0,12% ***	0,00% ***	-0,36% ***
ACCA	indirect	0,03% **	0,04% ***	0,00% **	-0,09% *
	direct	0,17% ***	0,07% **	0,01% ***	0,68% ***
	global	0,20%	0,12%	0,01%	0,59%
ACCAC	indirect	0,02% **	0,04% **	0,00% ***	-0,14% *
	direct	0,22% ***	0,14% **	0,01% ***	1,21% ***
	global	0,24%	0,18%	0,01%	1,07%
ACCRN	indirect	0,01% ***	0,03% ***	0,00% ***	-0,06% **
	direct	0,01% *	0,00% *	0,01% ***	-0,19% *
	global	0,03% *	0,04% *	0,01% ***	-0,25% *

		TE	TH	HPLUI	NGEL
Tués					
TUEFE	global	0,11% ***	0,22% ***	0,00% **	-0,07% *
TUERS	global	0,09% ***	0,25% ***	0,00% *	-0,03% *
TUEAGG	global	0,12% ***	0,15% ***	0,00% *	-0,53% **
TUEA	indirect	0,06% **	0,10% ***	0,00% **	-0,20% *
	direct	0,13% *	0,19% *	0,00% *	1,14% **
	global	0,18% *	0,29% *	0,00% *	0,94% *
TUEAC	indirect	0,04% **	0,07% **	0,00% ***	-0,23% *
	direct	0,03% *	0,40% **	0,00% *	1,83% **
	global	0,07% *	0,46% **	0,00% *	1,60% *
TUERN	indirect	0,01% ***	0,03% ***	0,00% ***	-0,05% **
	direct	0,10% **	0,13% **	0,01% ***	0,35% *
	global	0,11% **	0,17% **	0,01% ***	0,30% *

Significativité des paramètres : ***(*t*-ratio supérieur à 2), **(*t*-ratio entre 1 et 2), *(*t*-ratio inférieur à 1)

Tableau 10.5 : Les paramètres du modèle complet sur données filtrées, sur 1975-1999

France entière, réseau secondaire, agglomérations																								
	AIC	MAPRE_R	R ²	Exogènes																				
				Cte	LGA	CFM	ICARB	PAAU	PAAC	PARN	DTE	DTH	DHPLUI	DNGEL	TE-S	TH-I	TH-S	HPLUI-S	NGEL-S	S1	S2	S3	VESADI	
Accidents																								
ACCFE	-1015,8	3,0%	71,0%	-0,042 ***	0,13 **	0,14 **						0,00081 ***	0,00040 **	0,00001 **	-0,00167 *	0,00057 *	-0,00936 ***	0,00089 *	0,00108 *	-0,00133 *	0,00366 **	-0,00301 ***	-0,00190 **	0,0034 ***
ACCRS	-818,783	4,02%	58,9%	-0,025 ***	0,07 *	0,17 ***	-0,09 *					0,00080 ***	0,00029 *	0,00000 *	-0,00235 *	-0,00230 *	-0,00625 **	0,00386 **	0,00006 *	-0,00170 *	-0,00057 *	-0,02169 ***	0,00496 **	
ACCAG	-922,984	3,44%	58,9%	-0,045 ***	0,28 ***	0,19 ***	0,08 *					0,00054 ***	0,00041 *	0,00001 **	-0,00324 **	0,00447 *	-0,00931 ***	0,00067 *	-0,00002 *	-0,00129 *	-0,00041 *	-0,02278 ***	0,00111 *	
Tués																								
TUEFE	-764,5	4,7%	65,5%	-0,037 ***	0,13 **	0,33 **	-0,05 *					0,00118 ***	0,00097 **	0,00002 *	0,00139 *	0,00585 *	-0,01220 ***	0,00207 *	0,00054 *	-0,00300 **	0,00750 ***	0,00291 **	0,00402 *	0,00972 ***
TUERS	-193,075	5,92%	52,3%	-0,017 ***	1,14 ***	0,05 *	0,29 *					0,00122 ***	0,00026 **	0,00003 *	-0,00371 *	-0,00076 *	-0,01929 **	0,01132 **	-0,003 **	0,00763 **	-0,02612 ***	-0,01722 **	0,01751 **	
TUEAG	-488,904	7,43%	58,9%	-0,064 ***	0,55 ***	-0,22 ***	0,34 ***					0,00116 ***	0,00086 *	-0,00001 *	-0,00040 *	0,07296 ***	-0,00994 **	0,00273 *	0,00326 *	-0,00591 **	0,00711 *	-0,00906 *	0,00770 **	
Autoroutes, autoroutes concédées et routes nationales																								
	AIC	MAPRE_F	R ²	Exogènes																				
				Cte	LGA	CFM	ICARB	PAAU	PAAC	PARN	DTE	DTH	DHPLUI	DNGEL	TE-S	TH-I	TH-S	HPLUI-S	NGEL-S	S1	S2	S3	VESADI	
Accidents																								
ACCA	-574,3	6,8%	54,6%	-0,022 ***				0,82 ***				0,00214 ***	0,00113 **	0,00013 ***	0,00723 ***	-0,00554 *	0,00532 **	0,00243 *	-0,00613 **	-0,00174 *				
ACCAC	-453,4	8,1%	47,2%	-0,046 ***					0,96 ***			0,00266 ***	0,00065 *	0,00011 ***	0,00795 **	-0,02594 **	0,00300 **	0,00676 **	-0,00509 **	-0,00091 *				
ACCRN	-784,0	4,9%	50,7%	-0,046 ***						0,49 ***		0,00039 **	-0,00021 *	0,00007 ***	-0,00139 *	0,00451 *	-0,00234 *	0,00001 *	0,00001 *	-0,00011 *				
Tués																								
TUEA	27,7	19,7%	42,3%	-0,089 ***				1,79 ***				0,00190 **	-0,00177 *	0,00020 ***	-0,00213 *	0,00072 *	-0,02817 **	0,02924 **	-0,04002 ***	0,01340 **				
TUEAC	175,2	25,8%	41,1%	-0,088 ***					1,63 ***			0,00030 *	-0,00079 *	0,00009 *	-0,01288 *	-0,07212 **	-0,02285 **	0,03428 **	-0,02028 **	0,03158 ***				
TUERN	-470,1	8,2%	51,1%	-0,041 ***						0,43 ***		0,00093 **	0,00043 *	0,00006 **	0,00274 *	-0,00302 *	-0,00621 **	0,00283 *	0,00325 *	-0,00126 *				
Parcours																								
PAAU	-1191,6	2,1%	57,0%	0,037 ***	0,41 ***	0,15 **	-0,13 ***					0,00014 *	0,00009 *	-0,00001 **	0,00038 *	0,00048 *	-0,00355 ***	0,00183 **	0,00039 *	-0,00218 **	-0,00137 *	0,00340 ***	0,00978 ***	-0,00017 *
PAAC	-1029,9	2,8%	59,0%	0,036 ***	0,52 ***	0,23 ***	-0,18 **					0,00021 **	0,00009 **	-0,00001 *	0,00098 *	-0,00211 *	-0,00317 **	0,00193 *	-0,00060 *	-0,00302 **	0,00321 *	0,00613 ***	0,01623 ***	0,00003 *
PARN	-1359,0	1,6%	58,0%	0,017 ***		0,11 ***	-0,07 ***					0,00028 ***	0,00022 **	-0,00001 **	0,00050 *	-0,00290 *	-0,00526 ***	0,00202 **	-0,00013 *	-0,00207 **	-0,00278 **	0,00184 **	0,00612 ***	-0,00021 *

Significativité des paramètres : ***(t-ratio supérieur à 2), *(t-ratio entre 1 et 2), *(t-ratio inférieur à 1)

Tableau 10.6 : Effets des variables conjoncturelles, sur 1975-1999

		DTE	DTH	DHPLUI	DNGEL	TE-S	TH-I	TH-S	HPLUI-S	NGEL-S	S1	S2	S3	VESADI
Accidents														
ACCFE	global	0,08% ***	0,04% **	0,001% **	-0,17%	0,06%	-0,94% ***	0,09%	0,11%	-0,13%	0,37% **	-0,30% ***	-0,19% **	0,34% ***
ACCRS	global	0,08% ***	0,03%	0,00%	-0,24%	-0,23%	-0,63% **	0,39% **	0,01%	-0,17%	-0,06%	-2,17% ***	0,50% **	
ACCAG	global	0,05% ***	0,04%	0,001% **	-0,32% **	0,45%	-0,93% ***	0,07%	0,00%	-0,13%	-0,04%	-2,28% ***	0,11%	
ACCA	indirect	0,01%	0,01%	-0,001% **	0,03%	0,04%	-0,29% ***	0,15% **	0,03%	-0,18% **	-0,11%	0,28% **	0,80% **	-0,01%
	direct	0,21% ***	0,11% **	0,01% ***	0,72% ***	-0,55%	0,53% **	0,24%	-0,61% **	-0,17%				
	global	0,23% ***	0,12% **	0,01% **	0,75% ***	-0,51%	0,24% **	0,39%	-0,58% **	-0,35%	-0,11%	0,28%	0,80%	-0,01%
ACCAC	indirect	0,02% **	0,01%	0,00%	0,09%	-0,20%	-0,31% **	0,19%	-0,06%	-0,29% **	0,31%	0,59% ***	1,56% ***	0,003%
	direct	0,27% ***	0,07%	0,01% ***	0,80%	-2,59% **	0,30% **	0,68% **	-0,51% **	-0,09%				
	global	0,29% **	0,07%	0,01%	0,89%	-2,80%	-0,01%	0,86%	-0,57% **	-0,38%	0,31%	0,59%	1,56%	0,003%
ACCRN	indirect	0,01% ***	0,01%	-0,0005% **	0,02%	-0,14%	-0,26% ***	0,10% **	-0,01%	-0,10% ***	-0,14% **	0,09% ***	0,30% ***	-0,01%
	direct	0,04% **	-0,02%	0,01% ***	-0,14%	0,45%	-0,23%	0,00%	0,00%	-0,01%				
	global	0,05% **	-0,01%	0,01% **	-0,11%	0,31%	-0,49%	0,10%	-0,01%	-0,11%	-0,14%	0,09%	0,30%	-0,01%

		DTE	DTH	DHPLUI	DNGEL	TE-S	TH-I	TH-S	HPLUI-S	NGEL-S	S1	S2	S3	VESADI
Tués														
TUEFE	global	0,12% ***	0,10% **	0,00%	0,14%	0,59%	-1,22% ***	0,21%	0,05%	-0,30% **	0,75% ***	0,29% **	0,40% **	0,97% ***
TUERS	global	0,12% **	0,03%	0,00%	-0,37%	-0,08%	-1,93% **	1,13% **	-0,30% **	0,76% **	-2,61% ***	-1,72%	1,75% **	
TUEAG	global	0,12% ***	0,09%	0,00%	-0,04%	7,30% ***	-0,99% **	0,27%	0,33%	-0,59% **	0,71%	-0,91%	0,77% **	
TUEA	indirect	0,03%	0,02%	-0,002% **	0,07%	0,09%	-0,64% ***	0,33%	0,07%	-0,39% **	-0,25%	0,61% ***	1,75% ***	-0,03%
	direct	0,19% **	-0,18%	0,02% ***	-0,21%	0,07%	-2,82% **	2,92% **	-4,00% ***	1,34% **				
	global	0,22% **	-0,16%	0,02% **	-0,14%	0,16%	-3,45% **	3,25% **	-3,93% **	0,95% **	-0,25%	0,61%	1,75%	-0,03%
TUEAC	indirect	0,03% **	0,01%	0,00%	0,16%	-0,34%	-0,52% **	0,31%	-0,10%	-0,49% **	0,52%	1,00% ***	2,65% ***	0,00%
	direct	0,03%	-0,08%	0,01%	-1,29%	-7,21% **	-2,29% **	3,43% **	-2,03% **	3,16% ***				
	global	0,06%	-0,06%	0,01%	-1,13%	-7,56%	-2,80% **	3,74%	-2,13%	2,67% **	0,52%	1,00%	2,65%	0,00%
TUERN	indirect	0,01% ***	0,01% **	-0,0004% **	0,02%	-0,12%	-0,23% ***	0,09% **	-0,01%	-0,09% ***	-0,12% **	0,08% ***	0,26% ***	-0,01%
	direct	0,09% **	0,04%	0,01% **	0,27%	-0,30%	-0,62% **	0,28%	0,33%	-0,13%				
	global	0,10% **	0,05% **	0,01% **	0,30%	-0,43%	-0,85% **	0,37%	0,32%	-0,21%	-0,12%	0,08%	0,26%	-0,01%

Significativité des paramètres : ***(*t*-ratio supérieur à 2), **(*t*-ratio entre 1 et 2), *(*t*-ratio inférieur à 1

Tableau 10.7 : Les données de sécurité routière de 1998

	Routes nationales	Autoroutes concédées	Autoroutes non conc.	Réseau national	Routes secondaires	Agglomérations	France entière
Trafic (10⁸ veh-km)	886,28	595,15	373,81	1855,24			
moy. mensuelle	73,86	50	31,15	154,6			
% France entière				environ 1/3			
Accidents	11 807	2 426	3 484	17 717	36 278	70 392	124 387
moy. mensuelle	984	202	290	1476	3023	5866	10366
% France entière	9,49%	1,95%	2,80%	14,24%	29,17%	56,59%	
Tués	1 928	341	130	2 399	4 373	1 665	8 437
moy. mensuelle	161	28	11	200	364	139	703
% France entière	22,85%	4,04%	1,54%	28,43%	51,83%	19,73%	
Longueur du réseau (km)	24000	6646	2117	32763			

Source : SETRA - SNRD (trafic) et BAAC (nombres d'accidents et de tués)

Tableau 10.8 : Liste des variables, unités, sources

Dénomination	Définition	Unité	Source
ACCFE	Nombre d'accidents France entière		BAAC
ACCRS	Nombre d'accidents sur le réseau secondaire		"
ACCAGG	Nombre d'accidents en agglomérations		"
ACCA	Nombre d'accidents sur l'ensemble des autoroutes		"
ACCAC	Nombre d'accidents sur les autoroutes concédées		"
ACCANC	Nombre d'accidents sur les autoroutes non concédées(*)		"
ACCRN	Nombre d'accidents sur les routes nationales		"
ACCRRN	Nombre d'accidents sur le réseau routier national(*)		"
TUEFE	Nombre de tués France entière		"
TUERS	Nombre de tués sur le réseau secondaire		"
TUEAGG	Nombre de tués en agglomérations		"
TUEA	Nombre de tués sur l'ensemble des autoroutes		"
TUEAC	Nombre de tués sur les autoroutes concédées		"
TUEANC	Nombre de tués sur les autoroutes non concédées(*)		"
TUERN	Nombre de tués sur les routes nationales		"
TUERRN	Nombre de tués sur le réseau routier national(*)		"
PAAU	Trafic sur les autoroutes	10 ⁸ véhicules-km	SNRD
PAANC	Trafic sur les autoroutes non concédées	10 ⁸ véhicules-km	"
PAAC	Trafic sur les autoroutes concédées	10 ⁸ véhicules-km	"
PARN	Trafic sur les routes nationales	10 ⁸ véhicules-km	"
PAERN	Trafic sur l'ensemble du réseau routier national	10 ⁸ véhicules-km	"
CFM	Consommation finale des ménages	10 ⁹ francs constants 1980	estim. Inrets
ICARB	Indice du prix des carburants voiture	base 100 en 1980	estim. Inrets
LGANC	Longueur des autoroutes non concédées	kilomètre	SETRA
LGAC	Longueur des autoroutes concédées	kilomètre	"
LGAU	Longueur des autoroutes	kilomètre	"
TE	Température en été (avril à septembre)	0,1° C	Météo France
TH	Température en hiver(octobre à mars)	0,1 °C	Météo France
HPLUI	Hauteur de pluie	mm	Météo France
NGEL	Nombre de jours de gel	jour	Météo France
ATYTES	Nombre de jours atypiques "supérieurs" pour la température en été	jour	Météo France
ATYTHI	Nombre de jours atypiques "inférieurs" pour la température en hiver	jour	Météo France
ATYTHS	Nombre de jours atypiques "supérieurs" pour la température en hiver	jour	Météo France
ATYHS	Nombre de jours atypiques "supérieurs" pour la hauteur de pluie	jour	Météo France
ATYNGELS	Nombre de jours atypiques "supérieurs" pour le nombre de jours de gel	jour	Météo France
S1, S2, S3	Nombre de jours témoins d'un effet calendaire exceptionnel, regroupés en 3 classes	jour	
VESADI	Nombre de vendredi-samedi-dimanche	jour	
(*) non modélisé directement			

Chapitre 11.

Applications des modèles espace état

Les références reprises dans ce chapitre, sont tirées de travaux faisant appel à la représentation markovienne minimale, réalisés en 1997 (Bergel, 1997) et réactualisés ici sur un jeu de données étendu, et de travaux faisant appel à la représentation structurelle, réalisés dans le cadre du projet européen Safetynet (Commandeur et al, 2007), (Bergel, Debbbarh, 2008) et (Stipdonk H.(Ed.), en cours de publication 2008). Des exemples d'application des modèles structurels à des jeux de données connus sont donnés dans (Commandeur, Koopman, 2007).

11.1. Objectif

Dans ce chapitre nous proposons des modélisations de type ARMA vectoriel. Nous nous intéressons à la représentation markovienne (ou espace état) minimale et à la représentation structurelle, qui fournit une forme explicite des composantes non observables : à savoir les niveau, pente et saisonnalité, d'un processus à tendance linéaire locale.

Afin d'illustrer l'intérêt de ces approches, nous donnons des applications qui s'inscrivent dans le cadre de la modélisation à trois niveaux du risque routier, considérée dans le chapitre 10.

Il s'agit comme précédemment mais dans un cadre élargi, de modéliser des indicateurs de risque définis pour différentes catégories de réseaux et pour différents niveaux de risque, en prenant en compte les deux mêmes facteurs de risque : l'exposition au risque et le facteur climatique. La période de calage est ici 1975-2000, et la base de temps le mois.

11.2. Problématique

Le premier objectif de l'approche vectorielle est *la prise en compte d'interactions* entre des phénomènes que l'on vise à modéliser - par exemple quand ils sont observés sur deux réseaux routiers, comme ici sur les routes nationales et sur les autoroutes. Cette approche permet de saisir des effets croisés entre deux variables définies sur ces deux réseaux, qui peuvent aussi être des effets communs cachés, induits par exemple par une variable exogène non prise en compte de manière explicite.

Comme indiqué dans le chapitre 3, tout processus ARMA vectoriel (ou VARMA) a une représentation markovienne, ou représentation espace état, équivalente à sa représentation

VARMA. Choisir de modéliser un processus endogène en faisant appel à une représentation markovienne nécessite d'estimer le vecteur d'état associé, vecteur non observable.

Néanmoins, même si faire appel à une représentation markovienne présente l'avantage de disposer d'une représentation simplifiée de la dynamique (le vecteur d'état est un autorégressif d'ordre 1), les composantes du vecteur d'état associé ne sont pas pour autant facilement interprétables. C'est la raison pour laquelle l'on s'intéresse à sa représentation par un modèle structurel, dans lequel l'état s'exprime en fonction des composantes non observables (niveau, pente et saisonnalité) du processus endogène à tendance linéaire locale.

La représentation espace état et la modélisation structurelle sont communément utilisées pour modéliser un processus endogène (voir pour le cas univarié Harvey, 1989). Nous en donnons ici des exemples d'utilisation pour un processus endogène vectoriel.

11.3. Démarche

Dans la section 11.4 nous proposons une modélisation espace état d'un vecteur endogène composé des nombres d'accidents enregistrés sur deux réseaux routiers : les routes nationales et les autoroutes, en prenant en compte deux facteurs de risque : l'exposition au risque et du facteur climatique. Nous déterminons la dimension de l'état et nous interprétons ses composantes en utilisant les valeurs propres de la matrice de transition F , et nous donnons une estimation des paramètres exogènes. Dans la section 11.5 nous proposons une modélisation structurelle, de type tendance linéaire locale avec saisonnalité, du même vecteur observé. Nous donnons les formes explicites des composantes de l'état et comme précédemment une estimation des mêmes paramètres exogènes. Dans la section 11.6, nous proposons une étude simultanée pour les trois niveaux du risque, en retenant comme précédemment une modélisation structurelle d'un vecteur observé à trois composantes : le volume de trafic, le nombre d'accidents et le nombre de décès, sur le réseau autoroutier concédé. Enfin, dans la section 11.7, nous donnons des tests de validation de la nature des composantes inobservables.

11.3.1. Base de données

Dans ce chapitre, nous actualisons la base de données mensuelle utilisée pour le modèle RES sur la période 1975-2000.

Les données observées sont : les nombres de véhicules-kilomètres, les nombres d'accidents corporels et les nombres de tués, enregistrés sur routes nationales et sur autoroutes concédées.

Les données exogènes utilisées pour mesurer l'exposition au risque sont le premier groupe de données définies ci-dessus, et les données exogènes utilisées pour mesurer le facteur climatique sont les variables relatives à la température, à l'occurrence de gel et à la hauteur de pluie déjà définies dans le chapitre 10. Rappelons que deux variables sont retenues pour mesurer la température, selon la période de l'année retenue : d'hiver pour octobre à mars et d'été pour avril à septembre.

11.3.2. Outils d'estimation

Nous utilisons le logiciel MODEST développé en langage IML sous SAS (Azencott et al, 1997c). Cet outil est dédié à la représentation markovienne minimale.

En effet, dans ce logiciel, le vecteur d'état correspond à une représentation particulière telle que la matrice de covariance instantanée du vecteur d'état soit diagonale, et telle que les deux perturbations des deux équations d'observation et d'état soient liées par une relation simple. Comme indiqué en 3.1.1, ces deux perturbations sont $W(t)$ et $K.W(t)$.

Dans ces conditions, la représentation est minimale et unique (ou parfaitement identifiable) dans la classe des représentations markoviennes. Ces conditions sont indispensables pour que l'on puisse estimer numériquement les paramètres du modèle.

Nous utilisons également le package Ox/SSfPack3.0, l'outil sans doute le plus approprié dans la littérature pour faire de la modélisation structurelle multivariée.

Nous donnons enfin en annexe d'autres résultats obtenus avec la procédure UCM sous SAS (version 9.1), procédure qui ne permet de traiter à ce jour que le cas univarié.

11.4. Représentation markovienne minimale (cas bivarié)

L'on vise ici à modéliser un vecteur des deux mêmes indicateurs définis sur routes nationales et autoroutes. Dans cette modélisation, nous nous intéressons à l'estimation des super-paramètres F , K et H . Nous déterminons le rang de la matrice F qui représente la dimension de l'état. Les informations sur les composantes de l'état sont obtenues par l'interprétation des valeurs propres de la matrice F . Une estimation des paramètres exogènes est également fournie.

11.4.1. Méthode

Nous utilisons la méthode, en trois étapes, décrite dans le chapitre 3, pour estimer les paramètres du système de la représentation espace état avec variables exogènes d'un processus Y_t donnée en 3.2.1:

- estimation vectorielle autorégressive afin d'évaluer grossièrement les effets exogènes,
- estimation fine de la dynamique de la série endogène corrigée des effets exogènes,
- ré-estimation des effets exogènes.

Nous donnons les équations d'observation et d'état pour un vecteur endogène de dimension $m = 2$ et un vecteur exogène de dimension $k = 6$

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = HX_t + \begin{bmatrix} \sum_{i=1}^k D_{1t} Z_{1,t-i} \\ \sum_{i=1}^k D_{2t} Z_{2,t-i} \end{bmatrix} + \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}$$

$$X_{t+1} = FX_t + K \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}$$

avec : $F(r,r)$, $K(r,2)$ et $H(2,r)$ trois matrices à coefficients réels,

$D(B)$ une matrice $(2,6)$ dont les termes sont des polynômes en B ,

r la dimension du vecteur X ,

et W_t une perturbation.

A titre d'exemple, nous retenons ici pour variables endogènes les logarithmes des nombres d'accidents corporels $AccRN_t$ et $AccAC_t$, définis respectivement sur routes nationales et sur autoroutes, soit :

$$Y_{2,t} = \text{Log}AccRN_t,$$

$$Y_{1,t} = \text{Log}AccAC_t.$$

Et pour variables exogènes les logarithmes des nombres de véhicule-kilomètres enregistrés sur chaque réseau, et quatre variables pour mesurer le facteur climatique, soit :

$$Z_{1,t} = \text{LogPaRN}_t$$

$$Z_{2,t} = \text{LogPaAC}_t$$

$Z_{i,t}, i = 3, \dots, 6$, les 4 variables climatiques définies dans le chapitre 10.

11.4.2. Résultats

Les résultats obtenus au cours des 3 étapes sont donnés dans le Tableau 11.1.

11.4.2.1. Estimation préliminaire des effets exogènes.

Une estimation préliminaire des effets exogènes est obtenue en utilisant la forme vectorielle autorégressive avec exogènes suivante:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} A_{11,1} & A_{12,1} \\ A_{21,1} & A_{22,1} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} A_{11,12} & A_{12,12} \\ A_{21,12} & A_{22,12} \end{bmatrix} \begin{bmatrix} Y_{1,t-12} \\ Y_{2,t-12} \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^6 D_{1t} Z_{1,t-i} \\ \sum_{i=1}^6 D_{2t} Z_{2,t-i} \end{bmatrix} + \begin{bmatrix} V_{1t} \\ V_{2t} \end{bmatrix}$$

avec : V_t . une perturbation.

11.4.2.2. Estimation de la dynamique du système .

Les valeurs propres de la matrice de transition F et celles de la matrice F-KH sont calculées.

Nous donnons ici une interprétation des valeurs propres de la matrice F, conformément à ce qui a été exposé en 3.3.1. Les composantes d'indices 1,2 et 13 représentent la tendance à long terme du vecteur d'état. Les autres composantes représentent la saisonnalité : en particulier les composantes d'indices 11, 12, 14 et 15 à long terme et les composantes d'indices 2 et 3 à court terme.

11.4.2.3. Estimation finale des effets exogènes

La ré-estimation des paramètres des exogènes est effectuée à F,K et H fixés à leur estimation déterminée à l'étape précédente.

Dans cette estimation finale, deux paramètres ont été contraints à zéro : nous avons ici fait l'hypothèse que le nombre d'accidents enregistré sur un réseau dépend du trafic effectué sur le même réseau, et ne dépend pas du trafic effectué sur l'autre réseau.

La validité du modèle est établie par les tests habituels.

S'agissant de l'interprétation des paramètres exogènes, l'élasticité apparente des nombres d'accidents au trafic est de l'ordre de 0,6 sur les routes nationales et de 0,8 sur les autoroutes concédées.

Quant aux effets directs des variables climatiques sur les nombres d'accidents (à niveau de trafic inchangé), au seuil de significativité habituel, la hauteur de pluie et la température d'hiver est significative des nombres d'accidents sur les routes nationales, alors que la hauteur de pluie, la température d'été et l'occurrence de gel sont significatives des nombres d'accidents sur les autoroutes concédées. La significativité et l'amplitude de ces effets climatiques directs sur les nombres d'accidents sont cohérents avec les résultats déjà obtenus précédemment. Nous en donnons dans la section suivante d'autres estimations.

11.5. Modélisation structurelle bimodale (cas bivarié)

L'on vise ici à estimer de manière stochastique chacune des composantes non observables des deux processus endogènes considérés précédemment, et nous considérons le vecteur endogène de dimension $m = 2$ déjà défini. Nous retenons pour chaque indicateur la forme générale d'un modèle à tendance linéaire locale et saisonnalité - auquel nous nous référons par la suite en le qualifiant de modèle structurel. Sous cette forme, le niveau, la pente et la saisonnalité sont stochastiques en plus de la variable endogène, ce qui porte à quatre le nombre total de perturbations à estimer par indicateur.

11.5.1. Méthode

Comme indiqué en 3.4, le modèle à tendance linéaire locale et saisonnalité est un cas particulier de la représentation markovienne, dans lequel les matrices F et H sont connues, et le rang de la matrice F (ou dimension de l'état) est fixé a priori. La matrice K est l'identité.

En reprenant ici la notation de Harvey (1989), le modèle structurel multivarié avec variables exogènes est donné sous forme matricielle par les deux équations matricielles définies en 3.4.3.

L'équation d'observation est donnée par :

$$y_t = \left(\begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \alpha_t + D(B)x_t + \varepsilon_t$$

$$\text{où : } y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, x_t = \begin{bmatrix} x_{1t} \\ \vdots \\ x_{6t} \end{bmatrix}$$

$$\text{et } \alpha_t = \left[\eta_{1,t}, \zeta_{1,t}, \omega_{1,t}, 0, \dots, 0, \eta_{2,t}, \zeta_{2,t}, \omega_{2,t}, 0, \dots, 0 \right]'$$

sont respectivement le vecteur de 2 variables endogènes. $y_{1,t}, y_{2,t}$ (ayant chacune pour niveau, pente et composante saisonnière $\mu_{j,t}, \nu_{j,t}, \gamma_{j,t}$, $j=1,2$, la perturbation de l'équation d'observation, un vecteur de 6 variables explicatives, et le vecteur d'état de dimension $2(s+2)$).

L'équation d'état est donnée par :

$$\alpha_t = \left(\begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & -1 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \alpha_{t-1} + \kappa_t$$

$$\text{où : } \alpha_t = \left[\mu_{1,t}, \nu_{1,t}, \gamma_{1,t}, \dots, \gamma_{1,t-s+1}, \mu_{2,t}, \nu_{2,t}, \gamma_{2,t}, \dots, \gamma_{2,t-s+1} \right]',$$

$$\text{et } \kappa_t = \left[\eta_{1,t}, \zeta_{1,t}, \omega_{1,t}, 0, \dots, 0, \eta_{2,t}, \zeta_{2,t}, \omega_{2,t}, 0, \dots, 0 \right]'$$

Ici \otimes désigne le produit de Kronecker et κ_t la perturbation de l'équation d'état.

Le modèle structurel multivarié avec variables exogènes peut être réécrit sous la forme scalaire suivante :

$$\begin{cases} y_{1,t} = \mu_{1,t} + \gamma_{1,t} + \sum_1^6 \lambda_{1,i} x_{i,t} + \varepsilon_{1,t} \\ y_{2,t} = \mu_{2,t} + \gamma_{2,t} + \sum_1^6 \lambda_{2,i} x_{i,t} + \varepsilon_{2,t} \end{cases}, \begin{cases} \mu_{1,t} = \mu_{1,t-1} + \beta_{1,t-1} + \xi_{1,t} \\ \mu_{2,t} = \mu_{2,t-1} + \beta_{2,t-1} + \xi_{2,t} \end{cases}$$

$$\begin{cases} \beta_{1,t} = \beta_{1,t-1} + \varsigma_{1,t} \\ \beta_{2,t} = \beta_{2,t-1} + \varsigma_{2,t} \end{cases}, \begin{cases} \gamma_{1,t} = - \sum_{i=1}^{s-1} \gamma_{1,t-i} + \omega_{1,t} \\ \gamma_{2,t} = - \sum_{i=1}^{s-1} \gamma_{2,t-i} + \omega_{2,t} \end{cases}.$$

Les matrices de variance-covariance des deux perturbations de l'équation d'observation et de l'équation d'état, ε_t et κ_t sont données par :

$$Var\varepsilon_t = \begin{bmatrix} Var\varepsilon_{1,t} & Cov(\varepsilon_{1,t}, \varepsilon_{2,t}) \\ Cov(\varepsilon_{2,t}, \varepsilon_{1,t}) & Var\varepsilon_{2,t} \end{bmatrix}$$

$$Var\kappa_t = \begin{bmatrix} B & 0 & 0 & 0 \\ 0 & C & 0 & 0 \\ 0 & 0 & D & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

avec :

$$B = \begin{bmatrix} Var\eta_{1,t} & Cov(\eta_{1,t}, \eta_{2,t}) \\ Cov(\eta_{2,t}, \eta_{1,t}) & Var\eta_{2,t} \end{bmatrix}$$

$$C = \begin{bmatrix} Var\xi_{1,t} & Cov(\xi_{1,t}, \xi_{2,t}) \\ Cov(\xi_{2,t}, \xi_{1,t}) & Var\xi_{2,t} \end{bmatrix}$$

$$D = \begin{bmatrix} Var\omega_{1,t} & Cov(\omega_{1,t}, \omega_{2,t}) \\ Cov(\omega_{2,t}, \omega_{1,t}) & Var\omega_{2,t} \end{bmatrix}$$

Nous reprenons ici l'exemple donné en 11.4.

11.5.2. Résultats

Nous donnons dans le tableau 11.2 les résultats obtenus sous Ox/SSfpack3.0, pour le vecteur des deux séries d'accidents définies sur routes nationales et sur autoroutes concédées.

Les tests d'indépendance, homoscedasticité et normalité ont été menés sur le résidu de chacune des deux équations et ils sont satisfaisants.

La significativité et l'amplitude des paramètres exogènes sont voisines de celles obtenues avec la représentation markovienne exposée dans la section précédente, à l'exception de l'élasticité apparente des nombres d'accidents au trafic sur autoroutes concédées dont l'estimation est ici plus élevée.

L'élasticité apparente des nombres d'accidents au trafic est de l'ordre de 0,7 sur les routes nationales et légèrement supérieure à 1 sur les autoroutes concédées.

Concernant les effets directs des variables climatiques sur les nombres d'accidents, au seuil de significativité habituel, seule la hauteur de pluie est significative des nombres d'accidents sur

les routes nationales, alors que la hauteur de pluie, la température d'été et l'occurrence de gel sont significatives des nombres d'accidents sur les autoroutes concédées.

Les interprétations des paramètres climatiques directs que nous donnons ici sont à rapprocher de celles qui ont été données en 10.4.3.

La température est liée de manière positive au nombre d'accidents sur autoroutes concédées : une augmentation d'un degré de la température d'été se traduit par une augmentation de l'ordre de 1,8% du nombre d'accidents.

La hauteur de pluie est liée de manière positive au nombre d'accidents sur routes nationales et sur les autoroutes concédées : 100 mm de hauteur de pluie se traduisent par une augmentation de 0,6% et de 0,83% des nombres d'accidents sur les routes nationales et sur les autoroutes concédées.

Enfin, l'occurrence de gel est liée aux nombres d'accidents de manière positive sur les autoroutes concédées. Ainsi, un jour de gel dans le mois se traduit par une augmentation 0.8% de nombres d'accidents sur autoroutes.

11.5.3. Apport de l'approche bivariée

Nous donnons dans les Tableaux 11.3 et 11.4 les résultats obtenus sous Ox/SSfpack3, par une modélisation univariée de chacune des deux séries d'accidents, considérée isolément.

Les résultats obtenus sont très voisins de ceux obtenus dans l'approche bivariée. L'interaction entre les deux séries d'accidents modélisées est prise en compte dans la structure des covariances des quatre perturbations, alors que dans le cadre univarié celles-ci sont considérées comme non corrélées.

11.6. Modélisation structurelle des trois niveaux du risque (cas trivarié)

Dans ce paragraphe, nous nous intéressons aux trois niveaux du risque routier : l'exposition au risque, le risque d'accident et le risque d'être tué dans un accident, en prenant en compte comme précédemment le seul facteur climatique, sur un réseau donné.

Les indicateurs retenus pour les deux niveaux de risque sont les nombres d'accidents et de tués. Nous retenons également dans un second temps les taux d'accidents et les taux de tués

L'effet des variables climatiques sur les nombres d'accidents et de décès est un effet global, dans la première approche, et un effet direct, à niveau de trafic inchangé, dans la seconde.

Nous considérons successivement les deux vecteurs suivants à trois composantes :

- les logarithmes des nombres de véhicules-kilomètres, des nombres d'accidents corporels et des nombres de tués,
- les logarithmes des nombres de véhicules-kilomètres, des nombres d'accidents corporels pour 100 millions de véhicules-kilomètres et des nombres de tués pour 100 millions de véhicules-kilomètres.

11.6.1. Méthode

Nous considérons d'abord le processus endogène qui a pour composantes les logarithmes des nombres de véhicules-kilomètres, d'accidents corporels et de tués, définis sur un réseau donné

$$y_{1,t} = \text{Log}V_t$$

$$y_{2,t} = \text{Log}A_t$$

$$y_{3,t} = \text{Log}F_t$$

avec : V_t le volume de trafic,

A_t le nombre d'accidents corporels,

et F_t le nombre de tués.

Le vecteur d'état, non observable, a pour dimension $3(s+2)$, et est constitué des niveaux, pentes et composantes saisonnières de chacune des trois variables $y_{1,t}$, $y_{2,t}$ et $y_{3,t}$.

Nous considérons ensuite pour vecteur endogène suivant :

$$y_{1,t} = \text{Log}V_t$$

$$y_{2,t} = \text{Log} \frac{A_t}{V_t}$$

$$y_{3,t} = \text{Log} \frac{F_t}{V_t}$$

11.6.2. Résultats

Nous donnons successivement dans les tableaux 11.6 et 11.7 les résultats obtenus pour les deux vecteurs définis précédemment sur le réseau autoroutier concédé.

Le Tableau 11.6 fait ressortir les effets des variables climatiques sur le trafic, les nombres d'accidents et les nombres de tués. Au seuil de significativité habituel, seule la hauteur de pluie est significative du volume de trafic ; seules la hauteur de pluie et la température d'été sont liées de manière significative au nombre d'accidents sur autoroutes concédées. Il s'agit là d'un effet climatique global sur les nombres d'accidents, puisque le trafic n'est pas pris en compte de manière explicite dans le modèle explicatif des nombres d'accidents.

Le tableau 11.7 fait ressortir les effets des variables climatiques sur le trafic, le risque d'accident et le risque d'être tué dans un accident. Au seuil de significativité habituel, comme précédemment, seule la hauteur de pluie est significative du volume de trafic. En revanche, la hauteur de pluie, la température d'été et l'occurrence de gel sont significatives du risque d'accident. Il s'agit là d'un effet climatique direct sur les nombres d'accidents, à niveau de trafic inchangé.

Ces résultats sont cohérents avec les autres résultats obtenus précédemment dans ce chapitre, par modélisation structurelle univariée et bivariée, déjà donnés dans le Tableau 11.5

Nous les rappelons pour le seul réseau autoroutier concédé.

Effet de la hauteur de pluie : 100 mm de hauteur de pluie se traduisent par une augmentation de l'ordre de 0,6% des nombres d'accidents sur les autoroutes (qui se décompose entre une augmentation de l'ordre de 0,8% d'effet direct, et une augmentation de l'ordre de 0,2% d'effet indirect - les deux effets étant significatifs).

Effet de la température d'été : une augmentation d'un degré de la température en été se traduit par une augmentation du nombre d'accidents de l'ordre de 2 % (qui se décompose entre une augmentation de l'ordre de 2,2% d'effet direct, et une augmentation de l'ordre de 0,2% d'effet indirect – seul l'effet direct ressortant de manière significative).

Enfin, l'occurrence de gel n'est pas significative au seuil habituel, seul son effet indirect l'est.

11.7. Dynamique des composantes inobservables

Des tests sur la dynamique de chacune des composantes peuvent être effectués de manière à affiner la modélisation : certaines composantes à nature déterministe sont à considérer comme telles, ce qui diminue le nombre de paramètres à estimer dans cette approche.

Pour confirmer les natures déterministes ou stochastiques des composantes de l'état, nous interprétons les résultats de tests disponibles dans les sorties de la procédure UCM sous SAS.

Il convient de souligner que dans le cadre de la modélisation structurelle univariée, on obtient les mêmes résultats des estimateurs des paramètres du modèle, que ce soit avec la procédure UCM sous SAS ou bien avec le package SsfPack 3.0 sous Ox.

Notamment, le tableau 11.8 fournit, des résultats des tests de Student, aux seuils 0,05 et 0,1, sur la dynamique des niveau, pente et saisonnalité. Dans le cas où l'effet dynamique d'une composante est rejeté selon le test de Student, on étudie sa significativité par le biais d'un test de Khi2.

Sur les deux réseaux principaux, les *routes nationales et les autoroutes*, les tests de Khi2 et les indices de AIC et BIC, nous mènent à retenir un modèle structurel à niveau dynamique et à pente et saisonnalité déterministes pour les séries d'accidents corporels.

11.8. Conclusion

Dans ce chapitre, nous nous sommes intéressée à deux types de modélisation vectorielle : la représentation markovienne (ou espace état) minimale, et la représentation structurelle d'un processus à tendance linéaire locale.

- Modélisation espace état par une représentation markovienne minimale : on détermine la dimension de l'état, qui correspond au rang optimum de la matrice de transition F du système. Il s'agit d'une représentation particulière du modèle le plus général, représentation qui a des propriétés intéressantes pour le calcul numérique des estimateurs - cf. la notion de "balanced models" donnée dans (Aoki, 1988) et dans (Desai et al, 1985). Dans ce cadre général, les composantes non directement observables du vecteur d'état n'ont pas une forme explicite, et on ne peut pas les interpréter facilement, même s'il existe une relation entre chaque valeur propre de la matrice F et chaque composante du vecteur d'état, comme indiqué en 3.3.1.

Cette représentation, contrairement au modèle structurel, n'impose aucune contrainte sur les paramètres du modèle espace état.

- Modélisation structurelle : en imposant des contraintes aux paramètres du modèle précédent, on estime les composantes inobservables, d'un processus à tendance linéaire locale.. L'avantage de ce type de modélisation est la forme explicite de l'état, constitué des niveau, pente et saisonnalité d'un processus ici corrigé d'effets exogènes.

Nous avons produit deux types de résultats :

- la dynamique des composantes inobservables d'un processus observé corrigé d'effets exogènes qui a été estimée, et assortie de tests de significativité sur la nature de ces composantes ;
- les estimations des paramètres exogènes, qui sont cohérentes avec des résultats déjà obtenus dans le chapitre 10.

11.9. Discussion et extensions en cours

Les points de discussion portent sur les deux apports des modèles structurels appliqués à des données de sécurité routière, comparativement à d'autres types de modèles. D'une part la connaissance de la structure à proprement parler (nature des composantes inobservables) dans un modèle à tendance linéaire locale pour le processus corrigé des effets exogènes, et d'autre part la prise en compte de facteurs de risque et de mesures réglementaires ou d'évènements liés à la sécurité routière.

Rappelons très brièvement les résultats précédemment obtenus avec des modèles structurels appliqués à des données de sécurité routière. L'exemple qui illustre l'apport de ce type de modèle, et le plus souvent cité est certainement la modélisation à tendance linéaire locale des nombres de conducteurs tués ou gravement blessés dans des accidents de la route au Royaume-Uni de Janvier 1969 à Décembre 1984 (Harvey, 1989). Dans cet exemple, le volume de trafic et le prix des carburants sont pris en compte de manière significative dans le modèle et l'impact de l'obligation du port de ceinture à l'avant instauré en février 1983 est démontré. Même si la publication des travaux de Harvey a donné lieu à un débat important sur les avantages comparés des modèles structurels et des modèles autorégressifs, qui penchait en faveur des premiers, Scott a confirmé avec des modèles autorégressifs établis sur les mêmes données l'évaluation de l'impact de la mesure produite par Harvey.

Plus récemment, une approche menée de manière comparative sur des données agrégées de tués pour quelques pays européens sur une base annuelle, en prenant en compte les données de trafic et les principales mesures de sécurité routière, a mis en évidence la nature des composantes des taux de tués observés dans chaque pays sur longue période (Lassarre, 2001), (COST 329, 2004).

Enfin, une approche à trois niveaux a été menée de manière comparative sur des données d'exposition au risque, d'accidents et de décès pour la France et les Pays-Bas sur une base trimestrielle et sur longue période, mais sans prise en compte de facteurs de risque (Commandeur et al, 2007).

La modélisation structurelle exposée dans ce chapitre vient compléter les résultats qui viennent d'être cités.

D'une part, dans l'application traitée ici, les nombres d'accidents corporels enregistrés sur routes nationales et autoroutes concédées ont un niveau stochastique, mais une pente et saisonnalité déterministes. Les résultats obtenus avec un modèle sans variables exogène sont les mêmes.

D'autre part, les résultats obtenus dans l'estimation des effets exogènes valident des approches menées avec d'autres types de modèles, tels que le modèle RES décrit dans le chapitre 10, et d'autres modèles de type ARMA (Dupont, Martensen (Eds.), 2007).

Les résultats obtenus sur routes nationales et autoroutes concédées montrent que l'effet du trafic est très significatif sur chaque réseau, avec des valeurs d'élasticité voisines de résultats connus par ailleurs. S'agissant de l'effet direct des variables climatiques sur les nombres d'accidents, on retiendra que sur les routes nationales seule la hauteur de pluie est significative, alors que sur les autoroutes concédées deux facteurs climatiques supplémentaires, la température d'été et l'occurrence de gel, sont significatifs. Les effets indirects de la hauteur de pluie sur les nombres d'accidents par le biais du trafic ressortent significatifs sur les deux réseaux.

Les extensions de ces travaux à d'autres réseaux et à d'autres niveaux du risque sont en cours. L'extension de ces modèles structurels des nombres mensuels d'accidents corporels est en cours pour d'autres réseaux, dans une approche multivariée (multi-réseaux) ou simplement univariée au niveau d'un pays. Ces modèles sont également utilisés sur des données de victimes (tués, blessés hospitalisés).

La modélisation des trois niveaux du risque intégrant les facteurs exogènes donnée ici sur le réseau autoroutier concédé est en cours d'extension à d'autres réseaux : les routes nationales en France, la région d'Athènes, la France entière et les Pays-Bas, de manière à valider les résultats obtenus sur des réseaux de nature comparable (Stipdonk(Ed.), SafetyNet Deliverable D7.10, en cours de publication 2008).

Tableau 11.1 : Modèle markovien minimal pour les nombres mensuels d'accidents sur les routes nationales et les autoroutes concédées, sur 1975-2000 (sorties de listing MODEST). .

Estimation préliminaire des effets exogènes (1^{ère} étape)

Equation : 1 Variable endogene : LACAC

Moyenne : 5.1182 Ecart-type : 0.3056

Ecart-type résiduel : 0.1074

	Coef.Estim.	Ec-type	T
LACAC_1	-0.072	0.0387	-1.855
LACAC_12	0.1773	0.0564	3.1419
LACCR_1	0.3689	0.0721	5.119
LACCR_12	0.2536	0.0721	3.518
HPLUI	608E-7	219E-7	2.7775
LPAAC	0.5657	0.0851	6.6479
LPARN	0.0966	0.1477	0.6542
NGEL	0.0118	0.003	3.8897
TE	0.0012	0.0003	3.6768
TH	0.0009	0.0004	2.3721
CTE	-2.456	0.4836	-5.078

Equation : 2 Variable endogene : LACCR

Moyenne : 7.2357 Ecart-type : 0.301

Ecart-type résiduel : 0.0649

	Coef.Estim.	Ec-type	T
LACAC_1	-0.113	0.0234	-4.849
LACAC_12	0.0299	0.0341	0.8766
LACCR_1	0.4462	0.0435	10.254
LACCR_12	0.447	0.0435	10.268
HPLUI	41E-6	132E-7	3.0991
LPAAC	-0.169	0.0514	-3.28
LPARN	0.4082	0.0892	4.5762

NGEL	-0.002	0.0018	-1.075
TE	0.0003	0.0002	1.4627
TH	0.0005	0.0002	2.2498
CTE	-0.012	0.292	-0.042

Estimation de la dynamique du système (2ème étape)

Valeurs propres de F

	Module	Argument
1	0.99	0
2	0.9058	0
3	0.9559	0.1685
4	0.9559	-0.169
5	0.9617	0.3339
6	0.9617	-0.334
7	0.9392	0.4992
8	0.9392	-0.499
9	0.9157	0.6759
10	0.9157	-0.676
11	0.7687	0.8417
12	0.7687	-0.842
13	0.7857	0
14	0.9519	0.8396
15	0.9519	-0.84

Valeurs propres de F-KH

	Module	Argument
1	0.8448	0
2	0.8985	0.1914
3	0.8985	-0.191
4	0.4884	0
5	0.9015	0.346
6	0.9015	-0.346
7	0.9132	0.499
8	0.9132	-0.499
9	0.8775	0.6886
10	0.8775	-0.689
11	0.6381	0
12	0.8233	0.8584

13 0.8233 -0.858
 14 0.8929 0.8484
 15 0.8929 -0.848

Estimation finale des effets exogènes (3ème étape)

Multiplicateurs dynamiques finaux (Impact des exogenes)

Retard : 0

Multiplicateur Ecart-type T

LACAC-HPLUI 0.0001 184E-7 4.2976
 LACAC-LPAAC 0.8374 0.023 36.38
 LACAC-NGEL 0.0137 0.0027 5.1182
 LACAC-TE 0.0014 0.0003 5.4915
 LACAC-TH 0.0014 0.0003 4.2204

Multiplicateur Ecart-type T

LACCR-HPLUI 0.0001 r 118E-7 5.6549
 LACCR-LPARN 0.5829 0.0497 11.728
 LACCR-NGEL 0.0002 0.0017 0.1438
 LACCR-TE 0.0001 0.0002 0.3506
 LACCR-TH 0.0005 0.0002 2.1291

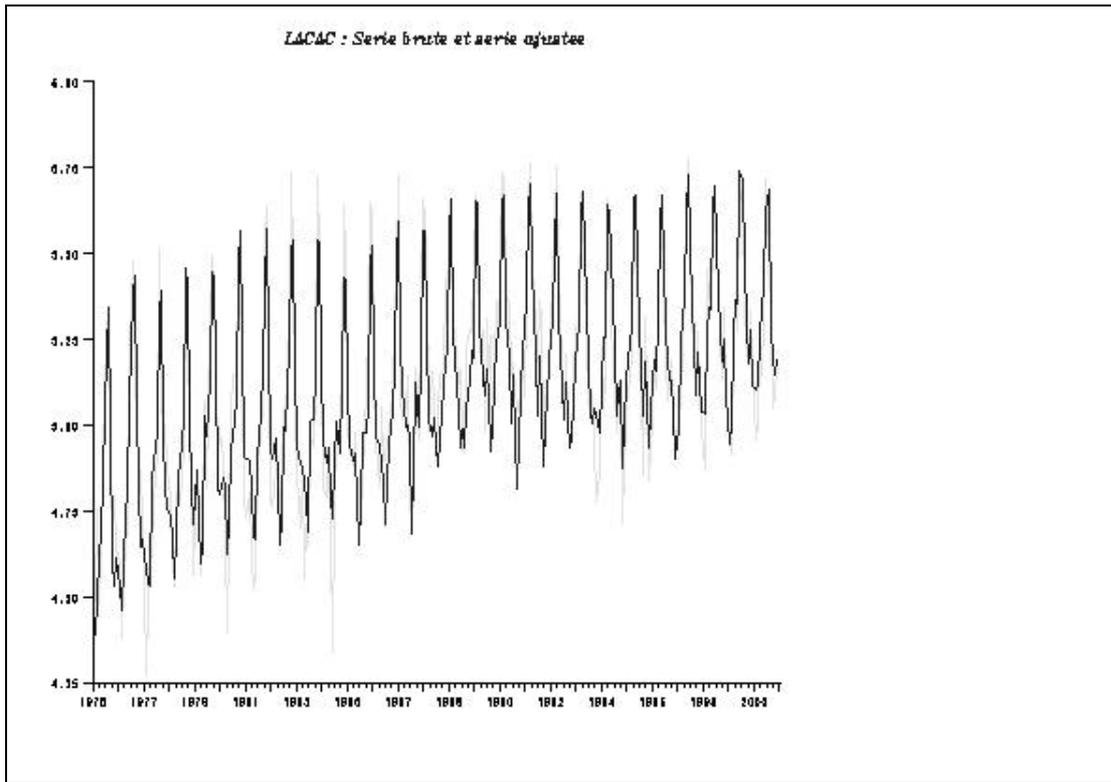


Figure 11.1 : Série brute et série ajustée des logarithmes des nombres mensuels d'accidents sur autoroutes concédées, sur 1975-2000 (sorties MODEST).

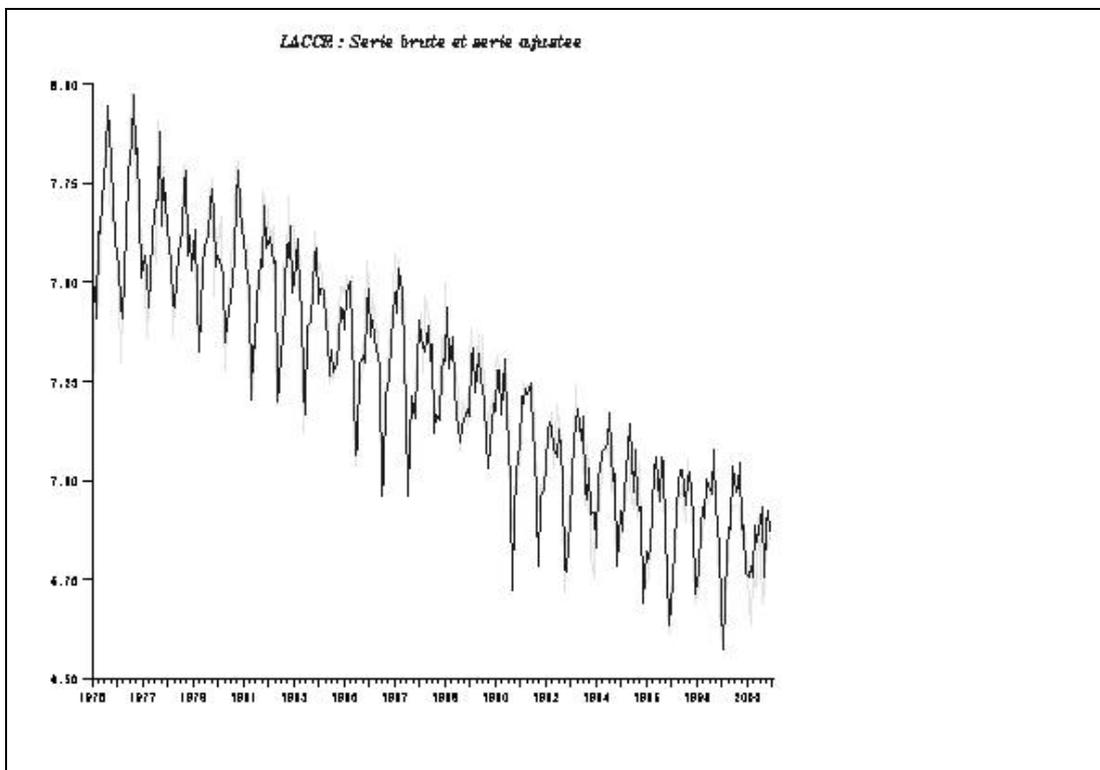


Figure 11.2: Série brute et série ajustée des logarithmes des nombres mensuels d'accidents sur routes nationales, sur 1975-2000 (sorties MODEST).

Tableau 11.2 : Modèle structurel bivarié pour les nombres mensuels d'accidents sur les routes nationales et les autoroutes concédées, sur 1975-2000 (sorties de listing Ox/ SSfPack).

Error variances

level error variance is 0.00045782 0.00033783 0.00033783 0.00029454	slope error variance is 1.4947e-008 -1.4586e-008 -1.4586e-008 1.4233e-008
seasonal error variance is 4.7888e-014 -1.5023e-014 -1.5023e-014 1.1304e-005	irregular variance is 0.0081593 8.7591e-005 8.7591e-005 0.0023825

Regression results

	Coefficient	standard error	t-value	p-value
LPaAC	1.037491256	0.1102404852	9.411163726	1.184779676e-018
	0	0	.NaN	1
	0	0	.NaN	1
LPaRN	0.686243412	0.1080312569	6.352267222	7.610449481e-010
TE	0.001853312958	0.0005609299407	3.304000773	0.001065540478
TE	0.000230899769	0.0003253153326	0.7097721684	0.4783824221
TH	0.0005134284949	0.000754913761	0.6801154269	0.4969422579
TH	4.952289444e-005	0.0004365130016	0.1134511326	0.9097468161
HPLUI	8.294029915e-005	2.07543839e-005	3.996278547	8.057163097e-005
HPLUI	6.044735416e-005	1.192575974e-005	5.068637597	6.924823879e-007
NGEL	0.008537924241	0.003953017068	2.159850083	0.03155564059
NGEL	-0.001950913105	0.002248314963	-0.867722333	0.3862219126

Tableau 11.3 : Modèle structurel univarié pour les nombres mensuels d'accidents sur les autoroutes concédées, sur 1975-2000 (sorties de listing Ox/ SSfPack).

Error variances

level error variance is	0.000259659
slope error variance is	1.14146e-013
seasonal error variance is	1.23103e-005
irregular variance is	0.00243525

Regression results

	coefficient	standard error	t-value	p-value
LPaRN	0.7255016033	0.1105703614	6.561447339	2.254179014e-010
TE	0.000229819246	0.0003292257156	0.6980598267	0.4856664479
TH	-2.779321459e-005	0.0004418748362	- 0.06289838732	0.9498882178
HPLUI	5.902625699e-005	1.206875664e-005	4.890831655	1.620271116e-006
NGEL	-0.002127146343	0.002275870709	- 0.9346516631	0.3507000279

Tableau 11.4 : Modèle structurel univarié pour les nombres mensuels d'accidents sur les routes nationales, sur 1975-2000 (sorties de listing Ox/ SSfPack).

Error variances

level error variance is	0.000179753
slope error variance is	4.66443e-014
seasonal error variance is	7.20193e-006
irregular variance is	0.00862065

Regression results

	coefficient	standard error	t-value	p-value
LPaAC	1.071145982	0.1118321538	9.578157495	3.429342755e-019
TE	0.001946378704	0.0005697451515	3.416226885	0.0007201896738
TH	0.0006664689277	0.0007661293898	0.8699169312	0.3850232457
HPLUI	7.737662955e-005	2.110258326e-005	3.666689931	0.0002893392222
NGEL	0.008342642349	0.004011431535	2.079716998	0.03837831974

Tableau 11.5 : Effets des variables climatiques pour 1975-2000 (sorties Ox/SSfPack)

		TE	TH	HPLUI	NGEL
LACCAC	direct	0,001946379 ***	0,000666469 *	7,74E-05 ***	0,008342642 **
	indirect	0,000121547 *	0,000327885 *	-1,81E-05 ***	-0,00232854 **
	total	0,002067925	0,000994354	5,93E-05	0,006014102
	global	0,001761827 ***	0,001029998 *	4,63E-05 ***	0,005090906 **
LACCRN	direct	0,000229819 *	-2,78E-05 ***	5,90E-05 ***	-0,002127146 *
	indirect	0,000140547 **	4,46E-04 ***	-7,18E-06 ***	-0,001065842 *
	total	0,000370367	4,19E-04	5,18E-05	-0,003192989
	global	8,49E-05 **	0,000302394 ***	4,81E-05 ***	-0,003136328 *

Significativité des paramètres : ***(*t*-ratio supérieur à 2), **(*t*-ratio entre 1 et 2), *(*t*-ratio inférieur à 1)

Nota : Ces estimations résultent des modèles univariés établis pour les nombres d'accidents :

- avec variables climatiques (effet global)
- avec trafic et variables climatiques (effet direct)

et des modèles univariés établis pour les nombres de véhicules-kilomètres avec variables climatiques (ce qui permet d'en déduire l'effet indirect).

Tableau 11.6 : Modèle structurel trivarié pour les nombres mensuels de véhicule-kilomètres, d'accidents et de tués sur les autoroutes concédées, sur 1975-2000 (sorties de listing Ox/ SSfPack).

Error variances

level variance is	slope variance is
1.4615e-018 -4.0828e-019 -7.6215e-019	3.2751e-007 2.6120e-007 5.5988e-007
-4.0828e-019 0.00017097 0.00011149	2.6120e-007 2.0832e-007 4.4653e-007
-7.6215e-019 0.00011149 7.2697e-005	5.5988e-007 4.4653e-007 9.5712e-007
seasonal variance is	irregular variance is
0.00048199 0.00063387 0.00079030	0.00061960 0.00037594 -9.6986e-005
0.00063387 0.00085014 0.0010992	0.00037594 0.0085160 0.0085981
0.00079030 0.0010992 0.0015124	-9.6986e-005 0.0085981 0.084323

Regression results

	coefficient	standard error	t-value	p-value
TE	0.0001475956446	0.0002118415051	0.6967267557	0.4864993046
TE	0.002226675471	0.0006002553184	3.70954726	0.0002462572817
TE	-0.0005402119212	0.001727442427	-0.3127235459	0.7547023387
TH	0.0002885033462	0.0002781915035	1.037067425	0.3005179325
TH	0.001071505228	0.0008025980028	1.335045969	0.1828472231
TH	0.002285129079	0.002317005996	0.986242195	0.3247883926
HPLUI	-1.678701187e-005	7.646259008e-006	-2.195454255	0.02887676356
HPLUI	6.337878404e-005	2.186618459e-005	2.898483902	0.004018895569
HPLUI	-2.360170096e-005	6.369238742e-005	-0.370557643	0.7112217479
NGEL	-0.00215651311	0.001421028625	-1.5175719	0.1301480715
NGEL	0.006612610941	0.004160182186	1.589500326	0.1129733924
NGEL	0.005546319911	0.01208402831	0.4589793874	0.6465727033

Tableau 11.7 : Modèle structurel trivarié pour les nombres mensuels de véhicule-kilomètres, d'accidents pour 100 millions de véhicule-kilomètres et de tués pour 100 millions de véhicule-kilomètres, sur 1975-2000 (sorties de listing Ox/ SSfPack).

Error variances

level variance is	slope variance is
1.0054e-009 -1.1505e-008 6.6522e-009	3.2750e-007 -6.6282e-008 2.9867e-007
-1.1505e-008 0.00017096 -5.9514e-005	-6.6282e-008 1.3415e-008 -6.0446e-008
6.6522e-009 -5.9514e-005 2.0725e-005	2.9867e-007 -6.0446e-008 2.7237e-007
seasonal variance is	irregular variance is
0.00048199 0.00015188 0.00015643	0.00061959 -0.00024361 -0.00047293
0.00015188 6.4390e-005 9.2592e-005	-0.00024361 0.0083837 0.00055530
0.00015643 9.2592e-005 0.00016417	-0.00047293 0.00055530 0.075644

Regression results

	coefficient	standard error	t-value	p-value
TE	0.0001475990032	0.0002118412249	0.696743532	0.4864888186
TE	0.002079093505	0.0005649651492	3.680038508	0.0002752132184
TE	-0.0027668794	0.001608967117	-1.719661869	0.08649850371
TH	0.000288499236	0.0002781911103	1.037054116	0.3005241247
TH	0.0007830061461	0.0007588064978	1.031891725	0.3029324291
TH	0.001213605333	0.002164666	0.5606432276	0.5754483743
HPLUI	-1.678724718e-005	7.646247843e-006	-2.195488234	0.02887430394
HPLUI	8.016545377e-005	2.062790861e-005	3.886261825	0.0001246771334
HPLUI	-8.698128056e-005	5.94794648e-005	-1.46237497	0.1446580862
NGEL	-0.002156560984	0.001421026538	-1.517607818	0.1301390134
NGEL	0.008769121869	0.003947608645	2.221375688	0.02705231958
NGEL	-0.001066233384	0.01130839288	-0.09428690667	0.9249425584

Tableau 11.8 : Test de significativité, sur 1975-2000 (sorties SAS)

	Level		Significativité		Slope		Significativité		Seasonal		Significativité	
	Est_var	Pr > t	Khi 2	Pr > Khi 2	Est_var	Pr > t	Khi 2	Pr > Khi 2	Est_var	Pr > t	Khi 2	Pr > Khi 2
LACCAC	0.00860	0.0669	7.02	0.0081	1.09919E-13	0.9988	22.39	<.0001	0.00000626	0.8413	32.87	0.0006
	0.00017908	0.0661	1.65	0.1988	0				0			
LACCRN	0.00031558	0.0768	54.64	<.0001	3.29195E-13	0.9988	14.54	0.0001	0.00001085	0.2737	318.73	<.0001
	0.00029148	0.1013			0				0			
	0				0				0			
LACCAC	BIC	AIC	R2	MAPE	Trafic		TE	P-value	TH	P-value	HPLUI	P-value
	-274.51448	-356.86055	0.85180	1.73099	1.06944	<.0001	0.00195	0.0006	0.00066705	0.3830	0.00007737	0.0002
	-285.95518	-360.81525	0.85171	1.73123	1.06906	<.0001	0.00194	0.0006	0.00065089	0.3951	0.00007799	0.0002
-249.91979	-321.03685	0.83319	1.89200	1.00814	<.0001	0.00225	0.0002	0.00088145	0.2876	0.00007087	0.0019	
LACCRN	-585.23009	-667.57616	0.95019	0.73418	0.71876	<.0001	0.00021012	0.5215	-0.00004290	0.9223	0.00005975	<.0001
	-594.14151	-669.00157	0.94982	0.73891	0.67828	<.0001	0.00070464	0.0007	0.00032151	0.2445	0.00003445	<.0001
	-510.36337	-581.48043	0.93470	0.83293	0.55035	<.0001	0.00013265	0.7374	0.00017188	0.7498	0.00004862	0.0010

Nota : La variable de trafic est LPaAC ou LPaRN, selon le réseau considéré

Figure 11.3: Série brute et série ajustée corrigée des exogènes des logarithmes des nombres mensuels d'accidents sur autoroutes concédées, sur 1975-2000 (sorties Ox/ SSfPack).

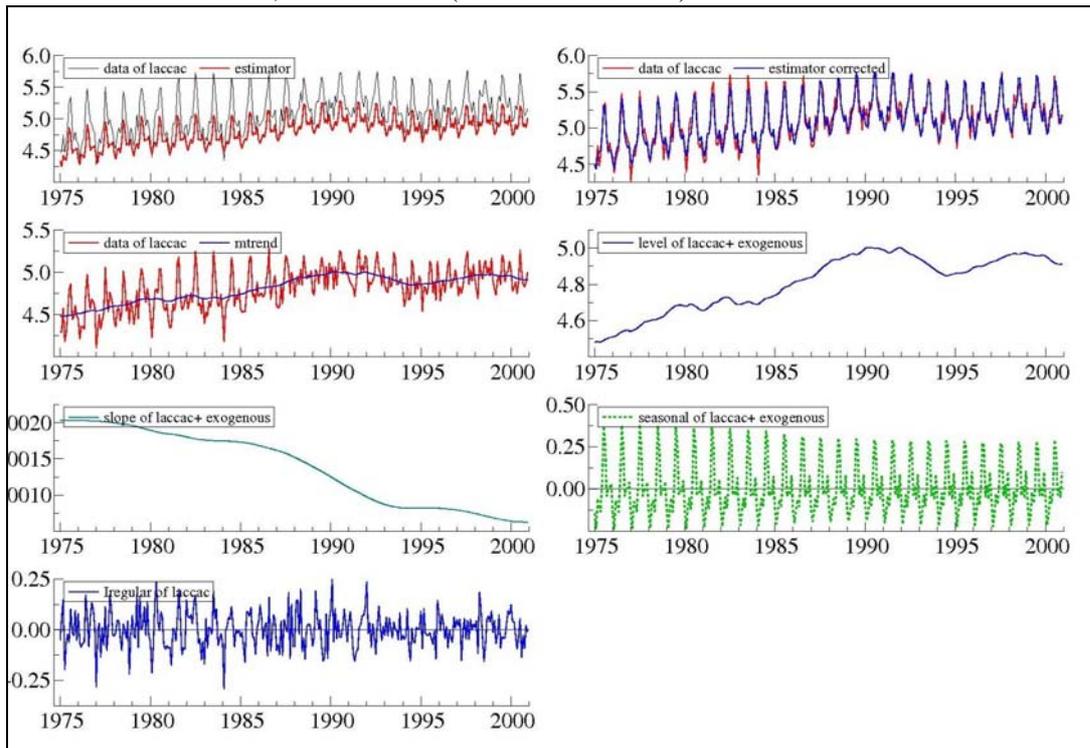
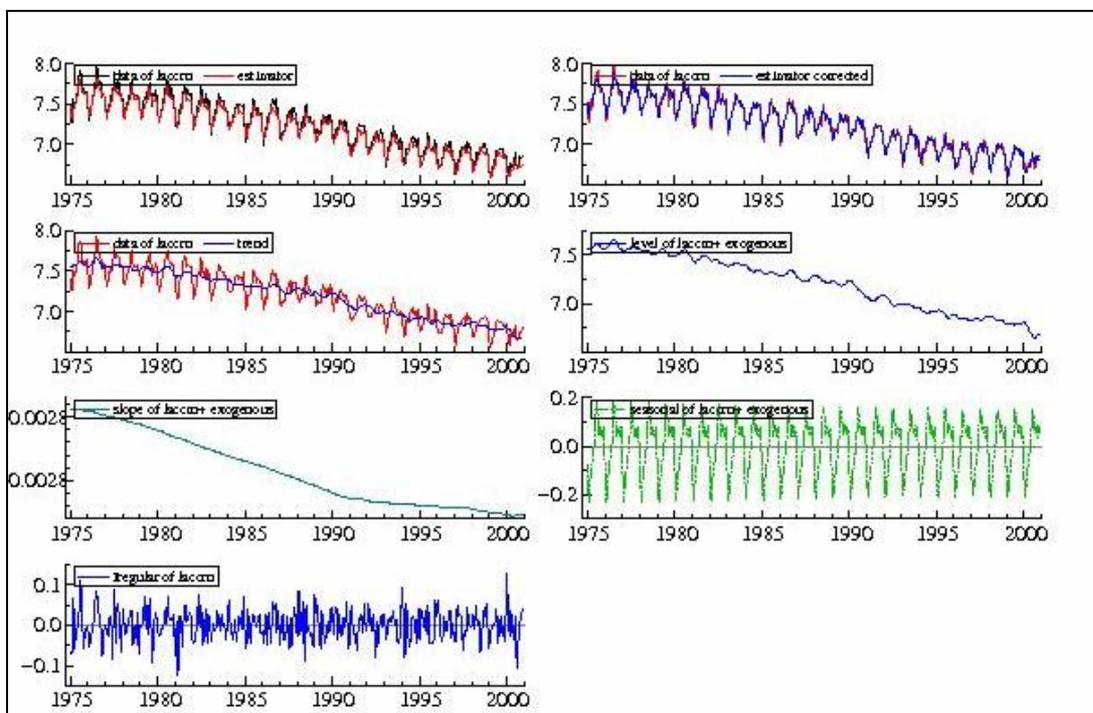
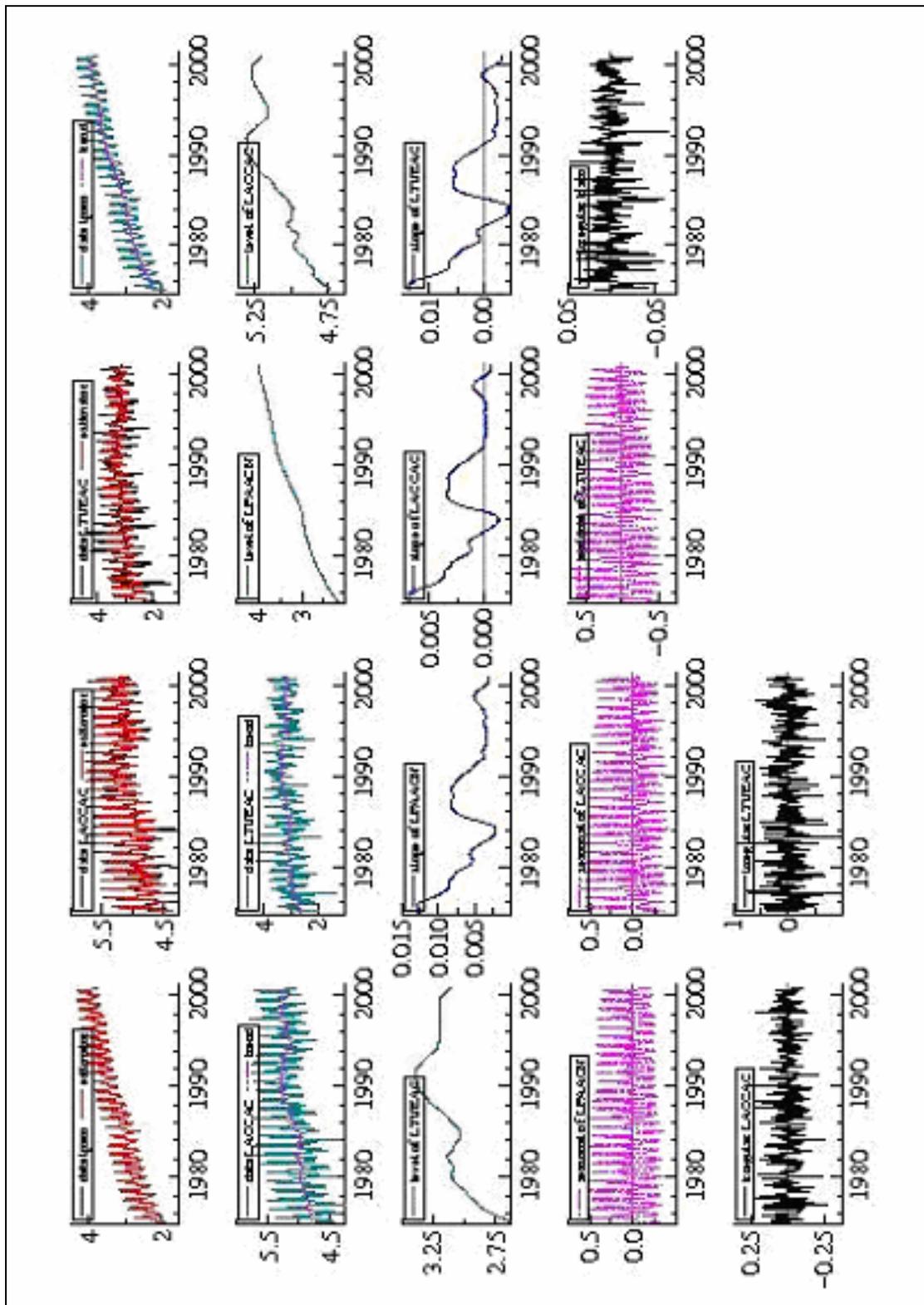


Figure 11.4: Série brute et série ajustée corrigée des exogènes des logarithmes des nombres mensuels d'accidents sur routes nationales sur 1975-2000 (sorties Ox/ SSfPack).



Nota: les niveau, pente et saisonnalité de la série ajustée corrigée des effets exogènes, sont également donnés.

Figure 11.5 : Séries brutes et séries ajustées corrigées des exogènes des logarithmes des nombres mensuels de véhicule-kms, d'accidents et de tués sur autoroutes concédées, sur 1975-2000 (sorties Ox/ SSfPack).



Nota : les niveau, pente et saisonnalité de la série ajustée corrigée des effets exogènes, sont également donnés.

Chapitre 12.

L'amnistie présidentielle

Nous présentons ici des travaux menés à la demande du Conseil National de Sécurité Routière, afin d'étudier les effets d'une annonce anticipée d'une amnistie présidentielle des infractions sur la sécurité routière, et qui ont donné lieu à un rapport d'expertise (Bergel et al, 2002). Des travaux menés en dehors de cette demande spécifique sont également discutés en fin de chapitre (Bourbonnais, Granger, 2002).

Il faut noter, comme le souligne d'emblée le groupe d'experts, que le recours à la modélisation statistique pour répondre à la question posée a été privilégié sans qu'une investigation approfondie sur les mécanismes de modification des comportements ait été menée. De plus, la réponse à la question posée est une éventuelle association statistique entre deux événements (l'annonce anticipée d'une amnistie présidentielle et des variations de l'accidentalité) et laisse entière la notion de causalité.

12.1. Objectif

Afin de préciser la relation entre la possible anticipation des effets de l'amnistie des fautes de conduite qui accompagne l'élection présidentielle et l'accidentalité, et en limitant l'analyse aux statistiques de décès - pour rendre compte des conséquences les plus graves des accidents de la route - et aux deux élections de 1988 et de 1995 - pour lesquelles l'information a été relayée par les médias - , un groupe d'experts réuni à la demande du Conseil National de Sécurité Routière a tenté de répondre à la question suivante :

« Peut-on mettre en évidence, en 1988 et/ou en 1995, dans les périodes encadrant les dates de l'élection présidentielle, une variation du nombre de tués sur les routes en France ? »

12.2. Méthode

Différentes approches par modélisation statistique étaient possibles, et trois types de modèles de séries chronologiques ont été retenues par le groupe⁷⁰: le modèle de désaisonnalisation Giboulée présenté dans le chapitre 10, le modèle de type ARMA présenté dans le chapitre 10

⁷⁰www.securite-sanitaire.org/route/rapportamnistie.htm

et un modèle GAM (Hastie, Tibshirani, 1990), qui est une extension des modèles GLM (McCullagh and Nelder, 1989, Dobson, 1990).

Nous exposons ici l'une de ces approches : l'analyse chronologique du nombre mensuel de tués, avec variables exogènes et variables d'intervention (ou modélisation ARIMAX) qui a été établie sur la période 1975-2001 avec le modèle à base mensuelle déjà cité.

Nous avons d'abord utilisé le modèle sur les données mensuelles brutes, puis utilisé une version simplifiée de ce modèle dans laquelle les facteurs transitoires de type climatique et calendaire ne sont pas pris en compte sur une base mensuelle comme les autres facteurs de risque, mais dans laquelle les données sont préfiltrées par une désaisonnalisation journalière effectuée avec le modèle Giboulée, La méthode adoptée est exposée plus avant dans les sections 1.2.2.1 et 1.2.2.2

12.2.1. Modélisation du nombre mensuel de tués, en données brutes

Le modèle prend en compte les variables exogènes suivantes : la consommation de carburants comme proxy (mesure) de l'exposition au risque (le kilométrage total effectué par les véhicules motorisés, qui mesure mieux l'exposition au risque, n'est pas estimé en rythme mensuel sur l'ensemble du territoire), le prix des carburants, et un petit nombre de variables météorologiques et calendaires qui rendent compte des effets transitoires, résultant des conditions climatiques et de la configuration calendaire.

La vitesse (vitesse moyenne, taux de dépassement de la vitesse limite, etc.) n'a pas été intégrée pour plusieurs raisons : elle n'est que partiellement disponible (seulement sur certains réseaux sur une période suffisamment longue) et d'autre part cette variable, traitée comme exogène, aurait masqué les effets de l'amnistie, que l'on cherche à mettre en évidence sur le nombre de tués et non pas sur les vitesses pratiquées. L'option de traiter la vitesse pratiquée comme une variable endogène subissant les effets de l'amnistie n'a pas été retenue, en raison de la non disponibilité de données de vitesse adéquates.

L'analyse d'interventions ne porte que sur la recherche d'éventuels effets des perspectives d'amnistie présidentielle de 1988 et de 1995 sur le nombre mensuel de décès. Il s'agit de déterminer une période d'action de l'amnistie sur le comportement des conducteurs et des policiers, puis d'identifier une forme d'intensité de cette action au moyen d'une fonction d'intervention. La forme de cette fonction est déterminée en fonction de l'allure des impacts

mensuels des variables indicatrices définies sur la période. On peut encore chercher, dans une dernière étape, à déterminer une période d'intervention optimale, en faisant varier les dates de début et de fin des périodes d'action supposées, de sorte à maximiser la vraisemblance du modèle.

12.2.1.1. Recherche des périodes d'intervention

La nature même de l'amnistie nous amène à délimiter la période d'action dans le temps (effet transitoire). Les deux périodes d'intervention retenues sont ici a priori fixées à novembre 1987⁷¹-juillet 1988 et septembre 1994-juillet 1995 (mois de la première annonce dans la presse - dernier mois précédant le vote de la loi d'amnistie).

Par ailleurs, les modèles utilisés font état de valeurs particulièrement faibles du nombre des tués, entre février 1987 et octobre 1987 : l'effet médiatique de l'affaire Anne Cellier (accident mortel d'une jeune femme, dans lequel le conducteur du véhicule responsable conduisait à très grande vitesse et en état alcoolique au moment de l'accident, et dont le procès s'est soldé par une peine d'emprisonnement de trois mois), puis à compter de juillet le vote d'une loi doublant les peines maximales pour conduite sous l'empire d'un état alcoolique ont certainement contribué à une prise de conscience de l'opinion et à une diminution de la gravité des accidents. En raison de sa proximité avec la période retenue pour l'amnistie de 1988, nous modéliserons l'effet de l'affaire Cellier et retenons la période d'avril à octobre 1987 comme période d'intervention supplémentaire. Là encore, la période retenue est limitée dans le temps dans la mesure où on admet que la loi aurait eu après son entrée en vigueur un effet limité sur l'aggravation de la peine effective prononcée.

Au total, les interventions à analyser sont au nombre de trois (Cellier, amnisties de 1988 et de 1995), et sont définies sur trois périodes retenues a priori.

⁷¹ On peut penser que l'effet Cellier se prolonge au-delà d'octobre 1987, auquel cas il se trouverait partiellement intégré dans l'effet de l'amnistie de 1988 tel que nous le modélisons, cet effet d'amnistie étant alors sous-évalué.

12.2.1.2. Recherche de la forme des fonctions d'intervention

La recherche de la forme des trois fonctions d'intervention a été réalisée à partir du modèle suivant :

$$(\Phi(B)(I - B12) \left[\log Y_t - \sum_{i=1}^I \alpha_i \text{Log} X_{i,t} - \sum_{j=1}^J \beta_j X_{j,t} - \sum_{k=1}^3 \sum_{l=0}^{n_k} \delta_{l,k} P^{T_{0,k}}(t-l) \right] = \mu + \Theta(B)a_t \quad (1)$$

avec :

Y le nombre de tués France entière,

$X_{i,i=1 \dots I}$ les I variables mesurant l'exposition au risque et les facteurs économiques,

$X_{j,j=1 \dots J}$ les J variables de nature météorologique et calendaire,

$P^{T_{0,k}}$, $k=1$ à 3 , trois variables indicatrices telles que $P^{T_{0,k}}(t)=1$ en $t=T_{0,k}$ et 0 sinon,

$T_{0,k}$ le premier mois de la k ème période d'intervention,

n_{k+1} le nombre de mois de la k ème période d'intervention

$\Phi(B)$ et $\Theta(B)$, deux polynômes de l'opérateur retard B ,

et a_t un bruit blanc.

La forme suggérée par le polynôme autorégressif $\sum_{l=0}^{n_k} \delta_{l,k} P^{T_{0,k}}(t-l)$ est un palier dans tous les cas (cf. annexes 11.1 à 11.3). Nous avons donc simplifié le modèle de départ (1) en utilisant trois variables indicatrices représentant des paliers.

$$\Phi(B)(I - B12) \left[\log Y_t - \sum_{i=1}^I \alpha_i \text{Log} X_{i,t} - \sum_{j=1}^J \beta_j X_{j,t} - \sum_{k=1}^3 \gamma_k \text{Step}_{k,t} \right] = \mu + \Theta(B)a_t \quad (2)$$

avec : $\text{Step}_{k,t}$, $k=1$ à 3 , trois indicatrices valant 1 en $[T_{0,k}, T_{0,k} + n_k]$, et 0 ailleurs.

12.2.1.3. Recherche des périodes d'intervention optimales

Nous avons encore ajusté le modèle (2) en faisant varier les dates de début et de fin des deux périodes d'intervention correspondant aux amnisties présidentielles, de manière à maximiser

la vraisemblance du modèle. Ceci nous a amenés à restreindre la deuxième période entre décembre 1994 et juin 1995, sans modifier la première.

12.2.2. Modélisation du nombre mensuel de tués, en données corrigées des variations locales au niveau du jour

La même démarche a été retenue pour modéliser les nombres mensuels de tués, corrigés des variations locales au niveau du jour – c'est-à-dire après prétraitement journalier des données, qui consiste à les corriger des effets locaux occasionnés par les conditions climatiques et calendaires. Le modèle mensuel des données ainsi préalablement corrigées en est simplifié :

$$\Phi(B)(I - B^{12}) \left[\log YC_t - \sum_{i=1}^I \alpha_i \text{Log} X_{i,t} - \sum_{k=1}^3 \gamma_k \text{Step}_{k,t} \right] = \mu + \Theta(B)a_t \quad (3),$$

avec : YC_t le nombre mensuel de tués France entière, corrigés des variations locales.

12.3. Résultats⁷²⁷³

12.3.1. Modèles sur les données brutes

Les résultats obtenus sont les suivants (Tableau 11.1). Succédant à un effet dû à l'affaire Anne Cellier de -6,5 % par mois (diminution moyenne de 6,5% du nombre de tués entre avril et octobre 1987), l'effet des perspectives d'amnistie de 1988 est estimé à +6,5% par mois (augmentation moyenne du nombre de tués de 6,5% entre novembre 1987 et juillet 1988), et celui de 1995 à +3,5% par mois (augmentation moyenne du nombre de tués de 3,5% par mois entre décembre 1994 et juin 1995).

En nombre absolu de décès, les effets des deux perspectives d'amnistie sont respectivement estimés à 517 et 168 tués. Les seuils de confiance associés sont respectivement de 0,04 et 0,28.

12.3.2. Modèles sur les données corrigées des variations locales au niveau du jour

Les résultats obtenus sur les trois périodes d'intervention sont légèrement moins significatifs que précédemment (Tableau 11.2). L'effet de l'affaire Anne Cellier est estimé à -7,6 % par mois d'avril à octobre 1987, l'effet des perspectives d'amnistie de 1988 à +4,6 % par mois de novembre 1987 à juillet 1988, et celui de 1995 à +2,3 % par mois de décembre 1994 à juin 1995.

En nombre absolu de décès, les effets des deux perspectives d'amnistie sont respectivement estimés à 368 et 112 tués. Les seuils de confiance associés sont respectivement de 0,15 et 0,47.

⁷² Rappelons que, dans ce qui suit, l'emploi du terme *effets* ne sous-entend pas qu'il y a nécessairement un lien de causalité entre les événements modélisés et la variation du nombre de tués estimée par modélisation, conformément à la remarque faite en introduction.

⁷³ Rappelons aussi qu'une investigation approfondie sur les mécanismes de modification des comportements des usagers de la route et des pratiques de contrôles et de sanctions dus à une annonce anticipée d'une potentielle amnistie présidentielle est nécessaire. Le rapport se limite en effet à mentionner les mécanismes pouvant théoriquement intervenir dans la modification des comportements des conducteurs à l'annonce d'une amnistie.

12.4. Conclusion

Les résultats obtenus à l'issue des deux modélisations entreprises, sur les nombres de décès en données brutes puis en données corrigées des variations locales au niveau du jour, vont dans le même sens, mais . les résultats obtenus avec les données corrigées des variations locales au niveau du jour sont très atténués et statistiquement moins significatifs.

L'ampleur de l'effet des perspectives d'amnistie de 1988 est plus grande (400 à 500 tués supplémentaires, de septembre 1987 à juillet 1988), qu'elle ne l'est pour celle de 1995 (100 à 180 tués de plus, de décembre 1994 à juin 1995). Toutefois, seul l'accroissement relatif à l'élection de 1988 est statistiquement significatif, au seuil habituel (soit 517 tués supplémentaires avec un seuil de confiance de 0,04).

Cette approche gagnerait à être affinée, car seules quelques variables explicatives ont été intégrées à l'analyse. Il aurait par exemple été souhaitable de relier la dégradation du nombre mensuel de décès à l'évolution des vitesses pratiquées, sur les deux périodes d'intérêt et sur les réseaux sur lesquels elles sont mesurées sur une période suffisamment longue. Ainsi, le modèle pourrait être ajusté sur les routes nationales et les autoroutes, sur lesquels le kilométrage mensuel, marquant l'exposition à la circulation et donc au risque d'accident, est mesuré, et sur lesquels l'effet des perspectives d'amnistie devrait être plus significatif.

12.5. Discussion

Un premier point de débat résulte de la comparaison entre les résultats des modèles ARIMA avec variables exogènes présentés dans ce chapitre : les effets évalués de l'amnistie sont différents, selon que les variations conjoncturelles (de type climatique et calendaire) sont prises en compte sur une base journalière ou mensuelle. Dans le premier cas, ils sont de l'ordre de 2/3 des effets évalués dans le second. Tout se passe comme si les facteurs transitoires, lorsqu'ils sont intégrés au niveau journalier, absorbent un tiers de l'effet amnistie évalué au niveau du mois. Ce constat vaut pour les deux amnisties.

Seul un critère de performance comparative entre ces deux approches permettrait de trancher entre ces deux résultats.

Un second point de débat résulte de l'absence de convergence entre toutes les approches statistiques menées sur les mêmes données, pour tenter de répondre à la même question posée.

Nous mentionnons ici les résultats obtenus par deux autres approches menées dans le même groupe d'experts sur une base journalière (Giboulée, et GAM), et par une approche menée en dehors de ce groupe sur une base mensuelle (Bourbonnais, Granger, 2002).

Le modèle Giboulée, qui est descriptif et ne permet pas d'analyser l'impact de facteurs particuliers, met en évidence des fluctuations de la tendance pendant les périodes précédant les élections présidentielles de 1988 et de 1995 (inflexion à la hausse modérée en 1995, et plus marquée en 1988 après la forte baisse de l'année 1987). Cependant, même si elles apparaissent visuellement marquées, ces évolutions de la tendance ne sont pas soumises à des tests de significativité.

Les résultats de la modélisation GAM, qui consiste en un ajustement non paramétrique avec forme linéaire des variables explicatives, sont inverses mais non significatifs : ils montrent une baisse non significative du nombre de tués en 1988 et en 1995, sur les périodes supposées d'effet d'amnistie. Dans cette modélisation, les facteurs de risque ne sont pas pris en compte de manière satisfaisante. De plus, les résultats sont incertains car le modèle différencie mal les effets de l'amnistie et les évolutions saisonnières normales du nombre de tués.

Enfin, l'approche menée à l'extérieur du groupe d'experts, par une modélisation essentiellement descriptive sans prise en compte de facteurs de risque, conclut à une augmentation significative, de l'ordre de 50 tués additionnels pendant les six mois précédant chacune des deux élections (soit de l'ordre de 300 tués additionnels sur les 6 mois).

En conclusion, nous pouvons insister sur l'importance du choix du pas de temps, et sur la nécessité d'un critère permettant de juger de l'adéquation du modèle aux données, qui n'a pas été utilisé ici.

De plus, le groupe de travail a conclu à une prise en compte insuffisante des facteurs de risque, et à la nécessité de mettre en œuvre un système statistique de recueil de données détaillées (indicateurs de sécurité routière et facteurs de risque à un niveau géographique fin ou à un niveau plus agrégé).

ANNEXES : Recherche de la forme des fonctions d'intervention

Graphique 11.1 à 11.3 : Impacts mensuels de l'effet Cellier (avril-octobre 1987), et des perspectives d'amnistie présidentielle de 1988 (novembre 1987-juillet 1988) et 1995 (septembre 1994-juillet 1995).

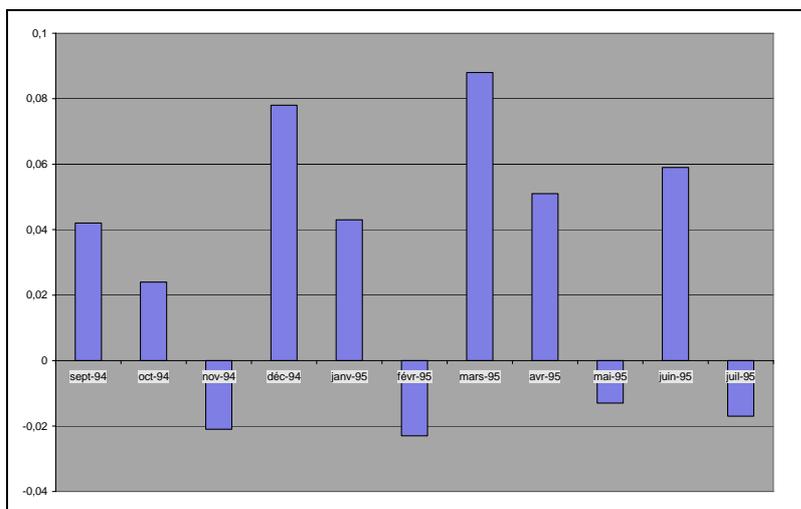
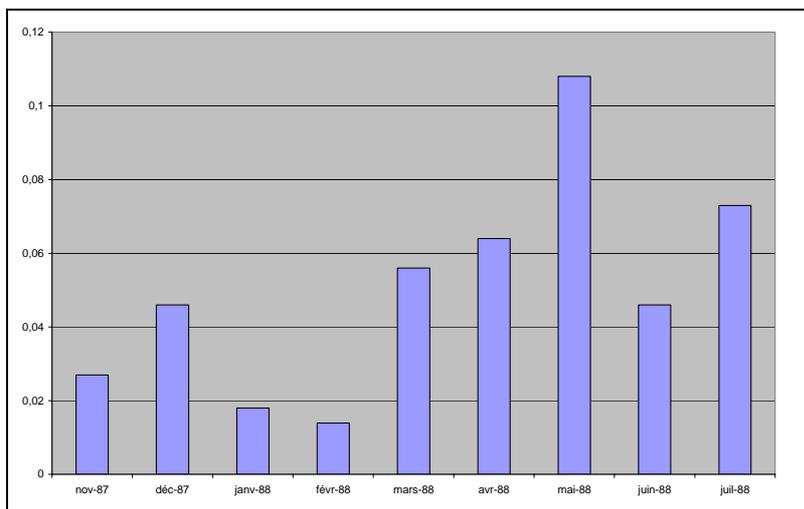
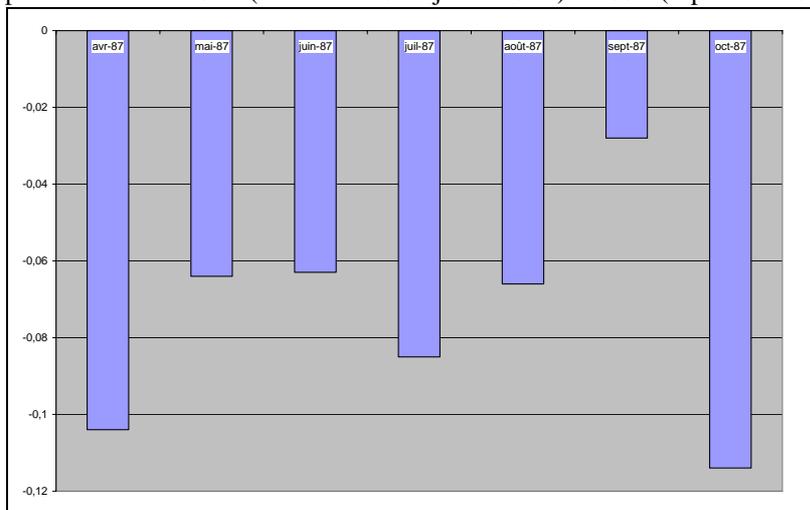


Tableau 11.1 : Le modèle sur les données mensuelles brutes, sur 1975-2001

Parameter	Estimate	Std Error	T Ratio	Lag	Variable	Shift
MU	-0.02452	0.0026660	-9.20		0 LTUEFE	0
MA1,1	0.77833	0.04043	19.25		12 LTUEFE	0
AR1,1	0.16357	0.05849	2.80		1 LTUEFE	0
AR1,2	0.14335	0.05910	2.43		2 LTUEFE	0
AR1,3	0.23408	0.05985	3.91		3 LTUEFE	0
NUM1	0.08976	0.07641	1.17		0 LCARBUB	0
NUM2	0.0013158	0.0003284	4.01		0 DTE	0
NUM3	0.0010321	0.0004655	2.22		0 DTH	0
NUM4	0.0000163	0.00001198	1.36		0 DHPLUI	0
NUM5	-0.0023010	0.0021175	-1.09		0 DNGEL	0
NUM6	-0.01055	0.0031936	-3.30		0 ATYTHI	0
NUM7	-0.0090827	0.0078523	-1.16		0 S1	0
NUM8	0.0044649	0.0044660	1.00		0 S3	0
NUM9	0.01387	0.0048195	2.88		0 DP	0
NUM10	0.0029882	0.0026133	1.14		0 VESADI	0
NUM11	-0.06497	0.03388	-1.92		0 STEP0	0
NUM12	0.06483	0.03206	2.02		0 STEP1	0
NUM13	0.03504	0.03238	1.08		0 STEP2	0
Constant	Estimate =	-0.0112535				
Variance	Estimate =	0.00403826				
Std Error	Estimate =	0.06354729				
AIC	=	-819.52059*				
SBC	=	-752.08894*				
Number of Residuals	=	313				

* Does not include log determinant

avec :

LTUEFE le logarithme du nombre mensuel de tués,

LCARBUB le logarithme de la consommation de carburants,

LICARB le logarithme du prix des carburants,

DTE-DTH la température maximale du jour, en été et en hiver, *DHPLUI* la hauteur de pluie et *DNGEL* l'occurrence de gel, moyennées sur une centaine de stations météorologiques réparties sur le territoire et sur le mois, mesurées par différence à la normale saisonnière,

ATYTHI le nombre de jours du mois où la température est particulièrement faible,

S1,S3 deux groupes de jours du mois atypiques au voisinage des jours fériés,

DP le nombre de jours du mois de grands déplacements liés à des départs et retours de congés,

VESADI le nombre de jours du mois de fin de semaine (vendredi, samedi ou dimanche),

STEP1 l'indicatrice de la période d'avril à novembre 1987,

STEP2 l'indicatrice de la période de novembre 1987 à juillet 1998,

STEP3 l'indicatrice de la période de décembre 1994 à juin 1995

(cf. la description précise des variables donnée dans l'annexe du chapitre 10).

Tableau 11.2. : Le modèle sur les données corrigées journallement des variations locales, sur 1985-2001

Parameter	Estimate	Std Error	T Ratio	Lag	Variable	Shift
MU	-0.03083	0.0031098	-9.91		0 LCVL	0
MA1,1	0.83096	0.04920	16.89		12 LCVL	0
AR1,1	0.20949	0.07795	2.69		1 LCVL	0
AR1,2	0.22803	0.07845	2.91		2 LCVL	0
AR1,3	0.09679	0.07852	1.23		3 LCVL	0
NUM1	0.24878	0.09135	2.72		0 LCARBUB	0
NUM2	-0.05583	0.07927	-0.70		0 LICARB	0
NUM3	-0.07620	0.03380	-2.25		0 STEP1	0
NUM4	0.04615	0.03181	1.45		0 STEP2	0
NUM5	0.02340	0.03227	0.73		0 STEP3	0
Constant	Estimate =	-0.0143571				
Variance	Estimate =	0.0035505				
Std Error	Estimate =	0.05958606				
AIC	=	-491.98961*				
SBC	=	-460.11575*				
Number of	Residuals=	179				

* Does not include log determinant

avec :

LCVL le logarithme des nombres de tués corrigés journallement des variations locales,

LCARBUB le logarithme de la consommation de carburants,

LICARB le logarithme du prix des carburants,

STEP1 l'indicatrice de la période d'avril à novembre 1987,

STEP2 l'indicatrice de la période de novembre 1987 à juillet 1998,

STEP3 l'indicatrice de la période de décembre 1994 à juin 1995

Conclusion et perspectives

Nous donnons d'abord les apports méthodologiques de la thèse. Les considérations de méthode sont exposées indépendamment des transports, puis de manière transversale aux deux champs d'application, en soulignant les caractéristiques des travaux menés avec le souci d'évaluer les modèles dans ces applications concrètes.

Nous donnons ensuite les principaux résultats obtenus et les perspectives par thème applicatif, pour les modèles de demande puis pour les modèles de risque routier.

Considérations méthodologiques

Considérations théoriques

Dans la première partie de la thèse, nous avons présenté une démarche méthodologique qui vise à prendre en compte, dans les modèles de séries temporelles, des effets exogènes mesurés à l'aide de variables additionnelles.

Dans le cadre de la thèse, ces variables additionnelles sont supposées déterministes. Toutefois, même si certaines d'entre elles étaient considérées comme aléatoires, elles seraient considérées comme causales : elles déterminent le processus endogène à modéliser, et ne sont pas déterminées par lui. En ce sens, nous les qualifions indifféremment *d'explicatives ou d'exogènes (exogenité forte)*.

Nous avons donné différentes représentations - ou spécifications économétriques - en insistant sur la structure du modèle associée à une représentation : ce point est important dans la mesure où il existe des équivalences entre certaines structures, et où le choix d'une représentation peut ne pas être exclusif d'une autre représentation. Nous nous sommes intéressée en particulier à deux types de structures : les modèles de décomposition d'une part, et les modèles autoprojectifs d'autre part (qui regroupent les autorégressifs, les ARMA et les ARIMA, et que nous avons qualifiés de modèles de type ARMA).

Nous avons détaillé la démarche retenue dans le cas d'une modélisation ARMA, puis dans le cas d'une modélisation de type VARMA, telle que la modélisation espace état et en particulier la modélisation structurelle.

Dans les cas où des variables exogènes ont été introduites dans la spécification, nous nous sommes intéressée à la dynamique d'un processus corrigé des effets des variables exogènes.

Les paramètres à estimer sont de deux sortes : d'une part les paramètres exogènes et d'autre part les paramètres de la dynamique (du processus corrigé des effets des variables exogènes).

Nous avons donné les méthodes d'estimation des modèles, considérés d'abord sans variables exogènes puis avec variables exogènes, en nous limitant au cadre linéaire - ou en nous y ramenant moyennant une transformation des données.

Les modèles ARMA ont été estimés par maximisation de la vraisemblance par rapport à l'ensemble des paramètres.

Les modèles multivariés ont dans un premier temps été traités par la représentation markovienne minimale, dans laquelle les variables exogènes interviennent dans l'équation d'observation. Un algorithme est utilisé qui propose le rang de la représentation markovienne minimale à partir du calcul de l'information mutuelle entre le passé et le futur définie par Desai et Pal (1985). Dans un deuxième temps, nous avons traité le cas des modèles structurels multivariés, dans lequel la dimension du vecteur d'état et les trois matrices du système sont connues; dans ce cas particulier, les modèles estimés sont à tendance linéaire locale. La dynamique a dans ce cas été estimée en utilisant le filtre de Kalman.

Considérations de type applicatif

Dans les deuxième et troisième parties de la thèse qui sont consacrées aux applications de cette démarche méthodologique aux transports, nous avons procédé comme suit. Nous avons d'abord rappelé dans les chapitres introductifs 4 et 8 le cadre méthodologique dans lequel s'inscrivent la modélisation de la demande multimodale de transport d'une part et la modélisation du risque routier d'autre part ; puis nous avons présenté dans les chapitres 5 à 7 et 9 à 12 un certain nombre d'applications dans chacun de ces deux champs.

Les modèles avec variables explicatives qui sont donnés ont été exploités différemment selon l'objectif fixé :

- *soit dans une visée explicative* : pour quantifier l'effet de facteurs connus par le biais de paramètres relatifs aux exogènes, comme pour l'évaluation de chocs exceptionnels et la prise en compte des facteurs de demande sur l'activité transport (chapitres 5, 6 et 7), pour la prise en compte des facteurs de risque et pour l'évaluation des effets de l'amnistie présidentielle sur l'accidentalité (chapitres 10 et 12) ;

- *soit dans une visée analytique* : pour mettre en évidence les mécanismes propres d'un processus après correction d'effets exogènes et les effets résiduels d'autres facteurs supposés, comme pour le prétraitement de l'effet météorologique local effectué dans le modèle Giboulée (chapitre 9) et pour l'estimation d'une tendance linéaire locale (chapitre 11).

Les effets pris en compte dans ces modèles ne font pas tous appel à des variables additionnelles, et ils ont été traités de différentes manières:

- *par filtre de différences* lorsque l'on ne dispose pas de variables pour les modéliser explicitement et qu'ils ne présentent pas un intérêt particulier : ainsi les composantes tendancielle et saisonnière d'un modèle ARIMA mensuel (chapitre 5), ou encore la composante hebdomadaire d'un modèle ARIMA journalier sont filtrées par un filtre de différences (chapitre 6) ;

- *par lissage par moyennes mobiles* lorsque l'on ne dispose pas de variables pour les modéliser explicitement, comme précédemment, mais qu'une estimation de certains effets est nécessaire : ainsi, certaines composantes tendancielle et calendaire d'un modèle de décomposition journalière (comme c'est le cas pour la méthode de correction des variations locales traitée dans le chapitre 9), et les composantes tendancielle et saisonnière d'un modèle de décomposition mensuelle (c'est le cas pour la méthode de correction des variations saisonnières traitée dans le chapitre 9) sont estimées par des techniques de lissage ;

- *par modélisation explicite* si l'on dispose de variables spécifiques pour mesurer des phénomènes observables, ou si l'on peut construire de telles variables ; nous en avons donné des exemples dans toutes les applications au secteur des transports présentées dans la thèse, qu'il s'agisse des applications à la demande multimodale de transport données dans les chapitres 5 à 7 ou des applications au risque routier données dans les chapitres 9 à 12.

Dans ce troisième cas, la prise en compte de variables additionnelles rend les modèles de séries temporelles "explicatifs" - sans que l'on puisse parler de causalité à proprement parler, et en notant que des corrélations entre les variables exogènes peuvent rendre difficile l'identification de l'effet propre à chaque variable. Le côté "boîte noire" des modèles vectoriels peut de plus rendre délicate l'interprétation de la dynamique du processus.

Ces modèles explicatifs permettent aussi un meilleur ajustement : meilleur suivi (utilisation du modèle sur le passé) et meilleure prévision (utilisation dans le futur) lorsque les variables exogènes peuvent être prévues, ou que l'on peut réaliser des scénarii réalistes pour anticiper leur évolution.

Les commentaires suivants portent sur le processus corrigé des effets exogènes, puis sur sa dynamique.

Les contraintes liées à la démarche résultent de la propriété de stationnarité des processus.

Dans tous les cas, pour utiliser des modèles ARMA, nous avons du supposer *la stationnarité d'un processus corrigé des effets exogènes* - ce qui s'est révélé pertinent dans les champs d'application retenus. En pratique, l'on introduit des filtres de différences sur les variables, endogène ou exogènes, de manière à contourner la contrainte de stationnarité sur le processus observé. De plus, la tendance et la saisonnalité peuvent être prises en compte par les variables exogènes de telle manière que le filtrage des données ne soit plus nécessaire.

Cette contrainte ne joue pas dans le cas de la représentation markovienne minimale et de la représentation par un modèle structurel.

S'agissant de la dynamique du processus corrigé des effets exogènes, les applications ont été menées avec un double objectif *de parcimonie et d'interprétabilité*.

Pour les modèles univariés, en raison d'écritures équivalentes, nous avons opté pour la forme ARMA de préférence à la forme autorégressive longue, du fait de son petit nombre de paramètres mais au prix d'une moindre qualité d'estimation des paramètres du polynôme moyenne mobile.

Pour les modèles multivariés, la recherche du rang de la représentation markovienne minimale permet de réduire le nombre de retards utiles pour prévoir le futur. Là aussi, la forme VAR, comme la forme autorégressive longue dans le cas univarié, n'a servi que d'écriture préliminaire pour obtenir une bonne estimation initiale de la matrice de variance-covariance du résidu.

Enfin, pour faciliter l'interprétabilité de la dynamique, nous avons opté pour une modélisation structurelle, dans laquelle on dispose d'une forme explicite de l'état, constitué des niveau, pente et saisonnalité.

Nous avons étudié, par des tests de stabilité, la robustesse de certaines applications par extension de la période de calage.

Enfin, l'évaluation des modèles utilisés dans les applications a été fondée sur les résultats des tests classiques d'hypothèse (relatifs au résidu du modèle et aux paramètres estimés), et sur

les critères habituels de validité statistique et de performance empirique (soit un critère d'information pénalisant et une mesure de l'erreur empirique).

Résultats et perspectives par thème applicatif

La modélisation de la demande de transport

L'objectif a consisté à prendre en compte les effets externes qui se manifestent dans le court terme sur les flux de trafics, que ce soit à l'horizon du jour, du mois, du trimestre ou du semestre.

La même structure de modèle a été retenue dans les applications à la demande de transport : il s'agit d'une structure ARIMA avec exogènes, et les variables additionnelles sont des variables indicatrices, des variables à valeurs discrètes ou réelles, discontinues ou continues.

Principaux résultats

Les modèles de demande présentés dans les chapitres 5 à 7 ont été utilisés pour l'analyse des évolutions infra annuelles (d'un jour à l'autre, d'un mois à l'autre, d'un trimestre à l'autre, d'un semestre à l'autre) d'indicateurs de trafics de voyageurs et de marchandises, agrégés par mode de transport et par catégorie de réseau.

Ces modèles ont été utilisés pour l'analyse sur la période de calage (modélisation descriptive et explicative) et pour la prévision dans le futur proche. Une prévision à quelques pas a dans certains cas été réalisée (extrapolation de tendance par ARIMA, ou prévision par ARIMA avec exogènes), mais sans que cet exercice se situe dans une optique de prévision à moyen/long terme.

Nous avons démontré que le gain obtenu avec de tels modèles, comparés aux mêmes modèles utilisés sans variable exogène, est important et que les paramètres relatifs à la partie exogène des modèles sont interprétables, même si dans le cas d'un groupe de variables exogènes corrélées, l'identification de l'effet relatif à chaque variable exogène peut être difficile.

Nous donnons ci-après des résultats résumés par chapitre.

Dans le chapitre 5, nous avons exposé une méthode d'*extrapolation de tendance par application de modèles ARIMA*, utilisée chaque trimestre pour anticiper à un horizon de six mois l'évolution de l'activité du secteur des transports.

Nous avons identifié, par une analyse systématique des écarts constatés entre les réalisations mensuelles et les projections réalisées le trimestre précédent, la nature et l'ampleur des principaux chocs qui ont affecté l'évolution des flux mensuels de cinq indicateurs modaux de trafic sur la période 1984-1991.

Les trois premiers portent sur les déplacements de voyageurs (la circulation interurbaine sur le réseau routier national, le trafic de voyageurs sur le réseau principal de la SNCF et le trafic de passagers d'Air Inter), et les deux derniers portent sur les trafics de fret (le transport routier de marchandises et le trafic de fret de la SNCF).

Les principaux chocs sont dus à une grève, à une météorologie atypique (vague de froid ou période de forte chaleur) et, de manière récurrente, à la configuration du calendrier.

Hormis les événements de nature exceptionnelle (grève, choc pétrolier, crise du Golfe, ...), les ordres de grandeur des impacts des chocs sur les trafics varient entre 5% et 10% de la valeur mensuelle attendue hors perturbation..

Des variables indicatrices ont été introduites dans les modèles ARIMA pour neutraliser ces perturbations dans les observations mensuelles de trafics, et ceci a permis d'améliorer la performance des modèles, mesurée par l'écart type résiduel sur la période de calage, qui diminue dans les mêmes proportions.

Dans le chapitre 6, nous avons considéré *les effets transitoires de nature climatique et calendaire sur les flux de trafic*.

Nous avons discuté les options liées à la méthode, selon que le pas de temps retenu pour l'analyse est le jour ou le mois. Un premier exemple porte sur le trafic de fret, ferroviaire et routier, sur 1975-1989. Nous avons estimé des poids journaliers, qui ont ensuite été exploités dans une analyse mensuelle. Un deuxième exemple porte sur la circulation interurbaine de véhicules légers sur routes nationales et sur autoroutes, sur 1985-1990. Nous avons estimé les « profils d'impact » du calendrier des jours fériés et du calendrier scolaire (calendrier exceptionnel) sur les flux journaliers de trafic. En utilisant directement un pas de temps mensuel, nous avons pris en compte simultanément les effets climatiques et calendaires sur la circulation interurbaine.

Le premier résultat est d'ordre statistique : le gain statistique est sensible avec un modèle ARIMA avec variables externes, puisque l'écart-type résiduel est réduit de 14,6% et de 35,2% pour les deux exemples des modèles mensuels de fret et de circulation interurbaine; le profil mensuel des séries se trouve mieux reproduit. Le second résultat est la quantification par un

paramètre de la relation entre chaque variable exogène et la variable modélisée - paramètre qui représente de fait un effet très moyen car agrégé. Enfin, on dispose de l'effet global des variables externes, mesuré par jour et par mois sur la période, et on peut ainsi mesurer les effets des deux groupes de variables, climatiques et calendaires, en les supposant indépendants.

Dans le chapitre 7, nous avons considéré *les effets à court terme des déterminants de l'offre et de la demande de transport sur les flux de trafic*.

Nous avons intégré les variables de mesure de l'offre (prix, consistance des réseaux) et de la demande (production/consommation), sur une base mensuelle et trimestrielle, dans des modèles explicatifs des trafics terrestres (route/fer) de voyageurs et de marchandises.

Des modèles ARIMA avec variables exogènes ont été estimés dans une approche modale, pour chacun des deux modes routier et ferroviaire, puis des modèles vectoriels autorégressifs avec variables exogènes ont été estimés dans une approche bimodale fer/route.

Pour étudier la cohérence de ces modèles avec des modèles annuels déjà établis sur les mêmes données, nous avons calculé des valeurs d'élasticité à un an et également retenu un critère de performance empirique à un an.

Deux exemples sont donnés, qui portent sur les trafics intérieurs de fret et sur les transports interurbains de voyageurs, pour la route et pour le fer, sur 1980-1992.

Le premier exemple fait apparaître une différence entre des effets de court terme et des effets de tendance qu'on mesure sur données annuelles : ainsi, l'effet du prix routier sur le fret routier est plus élevé à court terme qu'à un an. De plus, la recette unitaire ferroviaire (flux monétaire perçu à la tonne kilomètre transportée) est significative du fret routier, ce qui ne ressort pas des modèles à fréquence annuelle dans lesquels seul le prix routier est significatif.

Le deuxième exemple vise à expliquer le "choc" qu'a connu le trafic ferroviaire de voyageurs en début d'année 1993, au moment de la refonte du système tarifaire et de la mise en place du nouveau système de distribution "Socrate".

Enfin, la performance à un an de ces modèles est légèrement meilleure que celle des modèles annuels utilisés sur les mêmes données, et ceci vaut surtout dans l'approche bimodale. Pour autant, les modèles sont moins interprétables dans l'approche bimodale que dans l'approche modale, en raison du plus grand nombre de variables explicatives corrélées.

Une actualisation des modèles par extension de la période de calage, de 1980-1992 à 1980-1999, a été réalisée. Les modèles sont apparus robustes à un changement de format (passage

d'un format trimestriel à un format semestriel) et de logiciel (passage de SAS à E-views) ; en revanche, des changements de mesure de certains agrégats de la comptabilité nationale, constitués en nouvelle base 100 en 1995, ont induit des modifications notables dans les modèles antérieurs.

En résumé, il ressort des applications à la demande de transport que l'introduction de variables exogènes dans les modèles de type ARMA a produit deux types de résultats :

- rendre ces modèles explicatifs, dans la mesure où l'on dispose de paramètres pour interpréter la relation existant entre les facteurs de demande et l'indicateur d'intérêt, et où l'on parvient à quantifier l'effet global sur cet indicateur de groupes de variables, mesurés par jour et par mois sur la période,
- améliorer leur performance, le gain pouvant être d'un ordre de grandeur très significatif.

Perspectives

Cette approche menée à court terme sur des indicateurs agrégés de demande, peut être complétée en reconsidérant les deux niveaux, spatial et temporel, retenus dans la thèse. Au niveau spatial, une désagrégation géographique peut être menée, et au niveau temporel, les effets de long terme sont à étudier. Enfin, l'analyse des ruptures de tendance constatées sur la période récente reste à mener par une approche économétrique.

Désagrégation géographique : la prise en compte du zonage

La modélisation des trafics de marchandises est actuellement menée en désagrégant le territoire national en zones (Blardone, 2007).

Projections à 1-2 ans, Projections à moyen et long terme : les méthodes

Les estimations pour l'année en cours et l'année suivante (prévisions à 1-2 ans) les prévisions à moyen terme et long terme (5 ans, et 15-20 ans) sont les trois types de besoins récurrents pour anticiper l'évolution du volume d'activité transport.

Pour répondre à ces besoins, des approches complémentaires pourront être utilisées,

Les modèles économétriques utilisés sur données agrégées peuvent l'être sur une base annuelle, ou sur une base infra-annuelle par des méthodes de co-intégration.

La théorie de la co-intégration (Engle et Granger, 1987) vise à modéliser simultanément et distinctement des effets de court et de long terme : l'objectif est de quantifier des élasticités, à

court terme et à long terme, des trafics à leurs déterminants. Une analyse économétrique d'ensemble de la demande de transports terrestres de marchandises et de voyageurs a été produite sur données françaises (Lenormand, 2002). Cette direction de recherche apporterait un éclairage complémentaire aux travaux présentés dans la thèse qui n'ont été menés qu'en considérant le court terme.

Enfin, l'approche économétrique doit permettre de mettre en évidence, de quantifier et d'analyser les *ruptures de tendance* constatées dans le secteur des transports. De fait, les ruptures de tendance s'observent dans de nombreux marchés du transport (la circulation routière interurbaine, le transport ferroviaire de voyageurs, le trafic de fret ferroviaire), et dans d'autres secteurs que celui des transports. Les conséquences du renchérissement du prix des carburants sur le secteur des transports, parmi lesquelles l'inflexion spectaculaire à la baisse de la tendance de la circulation automobile interurbaine à compter de 2004, sont à analyser par une approche économétrique en réactualisant l'approche discutée dans le chapitre 7.

La modélisation du risque routier

L'objectif a consisté à prendre en compte les facteurs de risque qui se manifestent à court terme, en rythme journalier ou en rythme mensuel, sur le risque routier et notamment l'exposition au risque, le facteur climatique et le facteur calendaire.

Différentes structures de modèles ont été retenues dans les applications au risque routier : il s'agit de modèles de décomposition, à composantes déterministes ou stochastiques, et de modèles de type ARMA. Comme pour la demande de transport, les variables additionnelles sont des variables indicatrices, des variables à valeurs discrètes ou réelles, discontinues ou continues.

Principaux résultats

Les modèles de risque routier présentés dans les chapitres 9 à 12 ont été principalement développés à des fins de suivi conjoncturel (suivi descriptif, suivi explicatif), pour prendre en compte les principaux facteurs du risque routier tout en évaluant l'effet de mesures de sécurité routière ou d'événements liés à la sécurité routière.

Compte tenu de la part très importante de l'aléatoire dans le domaine de la sécurité routière, l'on s'est efforcé de neutraliser les fluctuations conjoncturelles de nature transitoire, afin de dégager une tendance : une tendance journalière a ainsi été estimée, puis une tendance mensuelle qui facilitent l'analyse de l'évolution des statistiques d'accidents corporels et de victimes de la sécurité routière (modèle Giboulée, dans le chapitre 9). Nous sommes parvenus

à prendre en compte les principaux facteurs de risque qui jouent à court/moyen terme dans un modèle à fréquence mensuelle, explicatif des nombres d'accidents corporels et de décès (modèle RES, dans le chapitre 10), et à confirmer les estimations des effets de deux facteurs de risque : l'exposition au risque et le facteur climatique, dans le cas où une autre forme de modèle à tendance linéaire locale est retenue (chapitre 11).

Nous avons enfin cherché à identifier et évaluer les effets d'annonce d'une amnistie présidentielle des infractions routières sur l'accidentalité, mesurée par le nombre total de décès sur la route (chapitre 12).

Nous donnons ci-après des résultats résumés par chapitre.

Dans le chapitre 9, nous avons exposé *la méthode de désaisonnalisation des indicateurs journaliers d'insécurité routière utilisée pour le modèle Giboulée*. Il s'agit d'une décomposition des données brutes, en composantes additives qualifiées d' "effets", qui est ici appliquée à des données d'accidents corporels et de victimes.

Une correction des variations locales est effectuée au niveau du jour, suivie d'une correction des variations saisonnières standard appliquée au niveau du mois. Les composantes journalières sont estimées par des techniques de lissage, à l'exception de deux d'entre elles qui sont estimées par appel à des variables exogènes : l'effet météorologique local et l'effet calendaire exceptionnel. L'effet météorologique local est estimé par un modèle de régression sur un grand nombre de variables météorologiques, et l'effet calendaire exceptionnel avec un modèle ARIMA avec variables indicatrices.

L'originalité de la composante météorologique locale tient au fait qu'elle a été établie à partir d'une information régionale, mesurée en 6 stations météorologiques avec des variables explicatives de l'influence du climat sur la fréquence et la gravité des accidents (ou clignotants). Ces clignotants diffèrent en effet selon l'indicateur modélisé, et selon la saison.

La composante calendaire exceptionnelle est estimée par appel à des variables auxiliaires (variables calendaires), codant des types de jours "exceptionnels" dans l'année.

La double correction de ces deux effets météorologique et calendaire constitue une correction locale originale, absente de la méthode standard de correction des variations saisonnières, et est ici adaptée au domaine de l'insécurité routière dans la mesure où les paramètres exogènes sont estimés par indicateur d'insécurité routière.

Les résultats obtenus pour la série du nombre de tués France entière sont donnés sur le deuxième quadrimestre 1992 pour les sorties journalières, et sur février 1985-décembre 1992 pour les sorties mensuelles.

Il ressort de l'analyse journalière que les fluctuations du nombre de tués sont sensiblement réduites après correction des effets locaux, et que la détection d'effets supplémentaires peut se faire à l'examen de la tendance : c'est le cas avec l'instauration du permis à points le 1^{er} juillet 1992, qui a eu pour effet d'infléchir à la baisse, à deux reprises, la tendance journalière du nombre de tués, alors que le profil moyen est à la progression à cette période de l'année.

En rythme mensuel, les effets locaux sont atténués. La série mensuelle corrigée des variations locales diffère encore de la série mensuelle brute, et leur écart mensuel peut dépasser les 10%, les mois de grand froid, et même atteindre 25% certains mois où les deux effets locaux se cumulent.

Dans le chapitre 10, nous avons discuté la manière de construire *un modèle explicatif du risque routier à petit nombre de facteurs de risque, significatifs sur une base mensuelle*. L'analyse porte sur l'évolution des nombres d'accidents et de décès, agrégés par catégorie de réseau ; à la différence de ce qui a été réalisé avec le modèle Giboulée, le pas de temps retenu est directement le mois.

Nous avons d'abord validé une forme fonctionnelle sur le réseau routier national (routes nationales et autoroutes), dans un modèle à deux niveaux de risque : le risque d'accident et le risque d'être tué dans un accident (mesurés par le nombre d'accidents et le nombre de tués), et à deux facteurs de risque : l'exposition au risque (mesurée par une variable principale de trafic) et le facteur climatique (mesuré par des variables secondaires relatives à la hauteur de pluie, à la température et à l'occurrence de gel). Différents tests de validation ont été effectués, sur deux périodes 1975-1993 et 1975-1998, qui produisent les mêmes conclusions. Il n'y a pas de différence significative entre le modèle avec transformation de Box-Cox sur l'exogène principale et le modèle avec transformation logarithmique sur l'exogène principale ($\lambda=0$), de sorte que l'on peut pour des raisons de parcimonie retenir cette seconde spécification, largement utilisée.

Nous avons ensuite étendu la spécification contrainte validée à l'étape précédente, aux trois niveaux de risque : l'exposition au risque, le risque d'accident et le risque d'être tué dans un accident (mesurés par le volume de trafic, le nombre d'accidents corporels et le nombre de tués) et à deux réseaux additionnels (le réseau secondaire et le milieu urbain) et à la France

entière, en augmentant le nombre de variables qui mesurent les facteurs de risque et en étendant la période de calage à 1975-1999.

Des liaisons significatives sont apparues entre les variables climatiques et les nombres d'accidents et de décès, pour la France entière et par catégorie de réseau. Sur les routes nationales et les autoroutes, l'effet global du climat a pu être séparé en deux composantes distinctes : son effet direct sur les nombres d'accidents et de tués, à volume de trafic inchangé, et son effet indirect par le biais du trafic.

Une analyse de l'effet météorologique sur les nombres mensuels d'accidents et de tués France entière est donnée, pour 1999 et pour 2000, qui ont été globalement défavorables à la sécurité routière. L'ordre de grandeur de l'effet météorologique est par exemple de 1%, en moyenne sur l'année 2000, de la valeur mensuelle de l'indicateur modélisé, qu'il s'agisse du nombre d'accidents ou du nombre de tués. Ce résultat est à mettre en perspective avec l'ordre de grandeur de l'effet mensuel qui dépasse les 10 % certains mois, et avec celui de l'effet journalier qui atteint des ordres de grandeur sensiblement plus élevés, tels qu'ils ressortent du modèle Giboulée.

Dans le chapitre 11, nous avons élargi le cadre précédent et considéré une *structure de modèle à composantes inobservables stochastiques* : il s'est agi comme précédemment mais sous l'hypothèse d'une tendance linéaire locale, de modéliser des indicateurs du risque routier définis pour différents niveaux de risque et catégories de réseaux, en prenant en compte les deux mêmes facteurs de risque : l'exposition au risque et le facteur climatique. La période de calage est 1975-2000, et la base de temps le mois.

Nous avons proposé des modélisations de type VARMA : nous nous sommes intéressée à la représentation markovienne (ou espace état) minimale afin de favoriser une parcimonie quant au nombre de paramètres à estimer et à la représentation structurelle afin de faciliter l'interprétation de l'état. L'avantage de la modélisation structurelle est la forme explicite de l'état, qui est constitué des niveau, pente et saisonnalité dans un modèle à tendance linéaire locale. Dans ce cas, les variances des perturbations associées au niveau, à la pente et à la composante saisonnière sont 3 paramètres supplémentaires à estimer, en sus de la variance de la perturbation associée au processus observé.

Une modélisation bivariée des nombres d'accidents sur routes nationales et sur autoroutes, confirme les résultats déjà obtenus dans l'estimation des effets exogènes dans le chapitre précédent. Des tests sont pratiqués relativement à la nature de leurs composantes, qui amènent

à conclure que les deux indicateurs est à niveau stochastique, et à pente (et saisonnalité) déterministes.

Enfin, une modélisation simultanée des indicateurs définis aux trois niveaux du risque est donnée, pour le réseau autoroutier concédé. Dans ce cas, les paramètres exogènes permettent d'estimer les effets climatiques directs sur les nombres d'accidents et les nombres de tués, à niveau de trafic inchangé, ainsi que les effets climatiques globaux sur les nombres d'accidents et les nombres de tués, résultant des variations simultanées du trafic.

Dans le chapitre 12, nous avons cherché à déterminer l'existence d'*un lien statistique entre l'annonce anticipée d'une amnistie présidentielle des infractions routières et l'accidentalité*. L'analyse a été limitée aux statistiques de décès et aux deux élections de 1988 et de 1995.

Nous avons exposé les résultats obtenus avec le modèle de type ARMA discuté dans le chapitre 10, avec lequel une analyse d'interventions a été menée. Il s'est agi de déterminer une période d'action de l'amnistie sur le comportement des conducteurs et des policiers, puis d'identifier une forme d'intensité de cette action au moyen d'une fonction d'intervention. Les périodes d'action retenues sont de 8 et 7 mois avant le vote de la loi d'amnistie, et la forme retenue pour la fonction traduit un effet constant au cours de la période.

Une version simplifiée du modèle, dans laquelle les effets transitoires de nature climatique et calendaire sont corrigés au niveau du jour, a aussi été utilisée.

Au vu des deux versions, l'ampleur de l'effet des perspectives d'amnistie de 1988 est plus grande qu'elle ne l'est pour celle de 1995. Seul l'accroissement relatif à l'élection de 1988 est statistiquement significatif, au seuil habituel (soit 517 tués supplémentaires avec un seuil de confiance de 0,04).

Un premier point de débat résulte de la comparaison entre ces résultats, qui diffèrent selon que les effets transitoires sont pris en compte au niveau du jour ou directement au niveau du mois.

Un second point de débat résulte de l'absence de convergence entre toutes les approches statistiques menées sur ces mêmes données, pour répondre à la même question. Au sein du groupe d'experts constitué pour apporter une réponse à l'administration des transports, deux autres approches par modélisation de séries chronologiques ont été menées qui ne débouchent pas sur des résultats significatifs. Une quatrième approche a été menée à l'extérieur du groupe d'experts, par une modélisation essentiellement descriptive sans prise en compte de facteurs

de risque et sur une base mensuelle, et conclut également à une augmentation significative, de l'ordre de 300 décès additionnels pendant les six mois précédant les deux élections.

Le choix du pas de temps, la performance intrinsèque des modèles et une meilleure prise en compte des facteurs de risque, ressortent comme les points à approfondir qui permettraient de faire évoluer la réponse faite ici à la question posée.

En résumé, il ressort des applications au risque routier que les mêmes facteurs de risque peuvent être pris en compte dans la modélisation du risque routier de manière très différente, selon le niveau de désagrégation spatio-temporel retenu, et que cette option a des conséquences sur l'évaluation d'autres paramètres liés à la sécurité routière lorsque ceux-ci sont estimés dans le même modèle.

Perspectives

Les perspectives consistent, là aussi, à reconsidérer les deux niveaux, spatial et temporel, retenus dans ces applications et à étendre de tels modèles aux données des Etats européens.

La désagrégation spatiale à proprement parler a été effectuée par catégorie de réseau, mais elle peut être menée suivant d'autres critères que le critère géographique, en retenant une typologie d'accidents et de victimes en fonction de différents critères.

Au niveau temporel, les effets de court terme mis en évidence sur une base journalière sont à approfondir, avec des modèles adaptés à de petits effectifs. Un modèle plus simple que le modèle Giboulée gagnerait à être exploité sur une base journalière, avec un petit nombre de variables climatiques à paramètres interprétables (Nallet, 2000 et Bergel, 2001), en retenant la démarche adoptée sur une base mensuelle présentée dans les chapitres 10 et 11.

Les effets de long terme restent à étudier, soit sur une base annuelle, soit sur une base mensuelle dans une démarche de co-intégration.

Les modèles économétriques développés sur des données agrégées au niveau national sont d'une part des modèles de la famille DRAG - ou s'y ramenant moyennant le choix d'une formulation simplifiée. Des modèles sont en cours de développement pour d'autres pays et par ailleurs un niveau de risque supplémentaire, lié au comportement, est retenu. Sur données françaises, par exemple, ce niveau est représenté par la vitesse pratiquée sur le réseau interurbain (Jaeger, 1998).

D'autres types de modèles économétriques ont été préconisés pour une analyse des données agrégées des états européens, dans le cadre du projet SafetyNet (<http://www.erso.eu>). Il s'agit de modèles de type ARMA et de modèles structurels (Dupont, Martensen (Eds.), 2007).

Des équivalences entre ces deux structures de base existent. Il est intéressant que les estimations des effets des mêmes facteurs de risque soient tout à fait cohérentes, avec ces deux structures. L'intérêt des modèles structurels est de fournir une représentation stochastique de la tendance, plus riche qu'une représentation déterministe. La prise en compte dans ce type de modèle de variables explicatives qui mesurent des facteurs de risque, qui n'est pratiquée à ce jour que dans le cadre univarié, est une direction de recherche à exploiter, en terme d'outils logiciels et en terme d'applications au champ du risque routier.

Pour autant, l'analyse par modélisation statistique de l'évolution de l'insécurité routière au niveau des états européens nécessite de disposer de bases de données de facteurs de risque, harmonisées entre les états.

Les orientations données portent d'abord sur la constitution d'une mesure d'exposition au risque (Haddak et al, 2005)(Yannis et al, 2005). En l'absence d'une mesure mensuelle de la circulation effectuée par les véhicules au niveau d'un pays, obtenue par enquête ou par sondage, une estimation du kilométrage mensuel par une modélisation économétrique s'appuyant sur la consommation mensuelle de carburants a été proposée pour quelques pays. Sur données françaises, la réactualisation de cette estimation effectuée sur 1957-1993 (Jaeger, 1998) est en cours sur 1957-2005.

Les facteurs de nature transitoire, climatique et calendaire, sont mesurés par des variables météorologiques ou par des variables à valeurs entières codant une typologie de jours. La constitution de telles bases de données ne pose pas de problème en soi. Un système d'information consistant à corriger les indicateurs de risque de l'influence de ces facteurs transitoires est à mettre en place au niveau des états européens, de manière à disposer d'un meilleur suivi conjoncturel du risque routier au niveau national. Avec un modèle à fréquence mensuelle, l'on disposerait d'une prévision à 3 à 6 mois (seules les différences entre les variables météorologiques et leur normale saisonnière seraient utilisées, qui sont nulles en prévision de sorte qu'une prévision des variables météorologiques n'est pas nécessaire), qui constituerait une référence dans le futur proche plutôt qu'une prévision réaliste. La base de temps pourrait également être le jour ou la semaine, avec un horizon de prévision qui se

situerait entre une semaine et un mois. L'intérêt d'un tel système d'information serait de s'appliquer à tout indicateur de risque, et par exemple simultanément aux trois indicateurs du risque routier définis sur un même territoire géographique, ou définis pour une même typologie d'accidents et de victimes.

La démarche méthodologique utilisée pour le risque routier s'applique à tout autre secteur de risque pour lequel l'on peut définir et mesurer à la fois: une exposition au risque, un risque et les conséquences de ce risque, et parmi lesquels la pollution, la santé, l'actuariat et la finance constituent des exemples de champs d'application à privilégier.

Annexes

Sigles utilisés

CEMS : Centre d'Etudes sur la Modélisation et les Statistiques

DAEI : Direction des affaires Economiques et Internationales

DRAST : Direction de la Recherche et des Affaires Scientifiques et Techniques

DSCR : Direction de la Sécurité et de la Circulation Routières

DTT : Direction des Transports Terrestres

INSEE : Institut National de la Statistique et des Etudes Economiques

OEST : Observatoire Economique et Statistique des Transports

ONISR : Observatoire National Interministériel de Sécurité Routière

SAE : Service d'Analyse Economique

SES : Service Economique et Statistique

SESP : Service des Etudes Economiques et de la Planification

SETRA : Service d'Etudes Techniques des Routes et des Autoroutes

SNCF : Société Nationale des Chemins de Fer

BAAC : Bulletin d'Analyse des Accidents Corporels

SNRD: Système National de Recueil de Données

Outils informatiques utilisés

Modélisation univariée de type ARMA :

Pour estimer les modèles univariés, nous avons principalement utilisé des procédures existantes sous SAS (procédures REG, AUTOREG et ARIMA pour estimer les modèles autoprojectifs ; procédures X11 et UCM pour estimer les modèles de décomposition, dans le cadre déterministe et dans le cadre stochastique.

Nous avons aussi utilisé un programme développé sous SAS qui permet de prendre en compte des variables exogènes dans les modèles de séries temporelles dans le cadre univarié (logiciel Mandrake, module ARMAX, Azencott et al, 1993).

Modélisation multivariée de type VARMA :

Pour estimer les modèles multivariés, nous avons utilisé un programme développé en langage IML sous SAS qui permet de prendre en compte des variables exogènes dans les modèles de séries temporelles dans le cadre multivarié, et en particulier pour la représentation markovienne minimale (logiciel MODEST, Azencott et al., 1997) ; un algorithme est utilisé qui propose le rang de la représentation markovienne minimale à partir du calcul de l'information mutuelle entre le passé et le futur définie par Desai et Pal (1985).

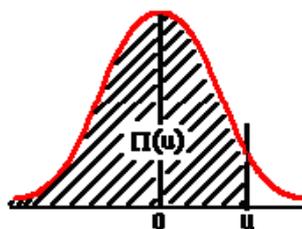
Nous avons aussi utilisé le package Ssfpack de Ox (<http://www.doornik.com/products.html>), qui permet de prendre en compte des variables exogènes dans les modèles structurels, dans le cas particulier où la dimension du vecteur d'état et les trois matrices du système sont connues ; dans ce cas les modèles estimés sont à tendance linéaire locale.

Table de Loi Normale

Fonction de répartition Π de la loi normale centrée réduite.

Probabilité de trouver une valeur inférieure à u .

$$\Pi(-u) = 1 - \Pi(u)$$

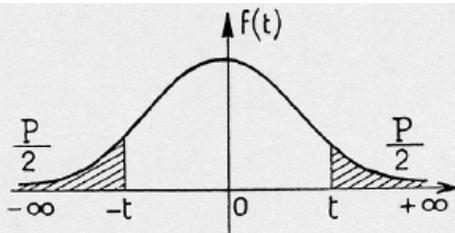


u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361

2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Table de distribution de T (Loi de Student)

Valeurs de T ayant la probabilité P d'être dépassées en valeur absolue.



$\begin{matrix} P \\ \backslash \\ v \end{matrix}$	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,617
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
80	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,650	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Table du Khi2*

La table donne la probabilité α pour que χ^2 égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.) ν .

$\nu \backslash \alpha$	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,016	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	1,064	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,467
5	1,610	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,515
6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	3,490	7,344	9,524	11,030	13,362	15,507	18,168	20,090	26,125
9	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	4,865	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	5,578	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	6,304	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	7,042	12,340	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	7,790	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	8,547	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	10,085	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,790
18	10,865	17,338	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	11,651	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,820
20	12,443	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,315
21	13,240	20,337	23,858	26,171	29,615	32,671	36,343	38,932	46,797

22	14,041	21,337	24,939	27,301	30,813	33,924	37,659	40,289	48,268
23	14,848	22,337	26,018	28,429	32,007	35,172	38,968	41,638	49,728
24	15,659	23,337	27,096	29,553	33,196	36,415	40,270	42,980	51,179
25	16,473	24,337	28,172	30,675	34,382	37,652	41,566	44,314	52,620
26	17,292	25,336	29,246	31,795	35,563	38,885	42,856	45,642	54,052
27	18,114	26,336	30,319	32,912	36,741	40,113	44,140	46,963	55,476
28	18,939	27,336	31,391	34,027	37,916	41,337	45,419	48,278	56,893
29	19,768	28,336	32,461	35,139	39,087	42,557	46,693	49,588	58,302
30	20,599	29,336	33,530	36,250	40,256	43,773	47,962	150,892	59,703

Exemple : avec $\nu = 1$ d.d.l. , pour $\chi^2 = 3,841$, la probabilité α est de **0,05**.

Quand le nombre de degrés de liberté est élevé, $\sqrt{2} * \chi^2$ est à peu près distribué normalement autour de $\sqrt{2} * \text{d.d.l.} - 1$ avec une variance égale à 1.

Bibliographie

AKAIKE H. (1974) Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics*, 26 : 363-387.

AOKI M. (1988) "State space models for vector-valued time series with random walk components" Paper presented at the 1988 ASA meeting.

AZENCOTT R, DACUNHA-CASTELLE D. (1984) *Séries d'observations irrégulières - Modélisation et prévision*, Masson.

AZENCOTT R., BERGEL R, FEVRIER E., HAUDEBOURG A, GARCIN M-L., LEUXE A., GIRARD B., MADRE J-L., NESPOUX V., POSTEL D., VILMART C.(1997a) *Modélisation bimodale des trafics terrestres de voyageurs (route/fer)*. Rapport du groupe de travail Beauvais Consultants/DIAM Recherche/DAEI/DTT/DRAST. Ministère de l'Équipement, La Défense.

AZENCOTT R., BERGEL R, FEVRIER E., HAUDEBOURG A, GARCIN M-L., LEUXE A., GIRARD B., NESPOUX V., VILMART C. (1997b) *Modélisation bimodale des trafics terrestres de marchandises (route/fer)*. Rapport du groupe de travail Beauvais Consultants/DIAM Recherche/DAEI/DTT/DRAST. Ministère de l'Équipement, La Défense.

AZENCOTT R., DURAND B., GIRARD B., GIRARD Y., VERNIER C. (1997c) *Un logiciel de modélisation de séries temporelles*. DIAM-Recherche/SAMOS/SCIPRE, Université Paris1-Panthéon-Sorbonne.

AZENCOTT R, GIRARD B, LETREMY P., ROY E. (1991) Constitution d'indicateurs météorologiques locaux adaptés à l'insécurité routière. Rapport OEST/CEMS. Paris: OEST.

AZENCOTT R, GIRARD B, LETREMY P., ROY E (1992) Modélisation ARIMA d'une dizaine d'indicateurs d'insécurité routière. Rapport OEST/CEMS. Paris: OEST.

BERGEL R. (1990) "*Séries temporelles : identifier les effets externes*". Notes de synthèse de l'OEST - 10/1990

BERGEL R. (1992) “*Utilisation de modèles ARIMA avec variables externes dans le secteur des transports*”. Proceedings of the 6th World Conference on Transport Research. Lyon

BERGEL R. (1997) *Multivariate road risk modelling on the French main road network* 25th European Transport Annual Meeting PTRC, Brunel University, Uxbridge, Londres

BERGEL R. (2001) “*A model of temporary effects on road traffic and road risk*” Actes de la Conférence PTRC, Cambridge.

BERGEL R. (2002) “*Modélisation des trafics de voyageurs et de marchandises – Formats trimestriel et semestriel – Agrégats macro-économiques base 1995 – Estimation sous SAS et sous E-views*”. Rapport de convention SES/INRETS n° 99/5815, Arcueil.

BERGEL R. (2003) *Modelling road and rail transport demand in the short : application and results*. European Transport Conference, Strasbourg.

BERGEL R., BLAIN J.C., JIANG F. (1995) Elasticités du trafics ferroviaire de voyageurs à la consommation et aux prix. Notes de synthèse de l’OEST n° 96 – 11/1995

BERGEL, R., CARNIS, L., LASSARRE, S., LE BRETON, P., LE TERTRE, A., PAGE, Y., THELOT B. & UHRY Z. (2002) *Amnistie présidentielle et sécurité routière*. Rapport d’un groupe d’experts appartenant à l’Inrets, à l’Invs, au Setra, au Lab, et à l’Université de Reims Champagne-Ardenne
<http://www.securite-sanitaire.org/route/rapportamnistie.htm>, <http://www.securite-routiere.org>

BERGEL R., CREBASSA P., PINON B. (1995) “*Un outil de suivi et de prévision à court terme du trafic aérien domestique*”. Notes de synthèse de l’OEST n° 97 – 12/1995

BERGEL R., DEBBARH M. (2008) *The three levels of the road risk process* Second SafetyNet Conference, Rome <http://www.erso.eu>

BERGEL R., DEPIRE A. (2000) *Modélisation multivariée d’indicateurs d’insécurité routière*. Rapport de convention Inrets/DSCR n° 98001, Arcueil

BERGEL R., DEPIRE A. (2002) *Modélisation du risque routier - Suivi explicatif et prévision – Mise en œuvre sur l'année 2000*. Rapport de convention Inrets/DSCR n° 01-70018, Arcueil

BERGEL R., DEPIRE A. (2004a) *Recherche d'une forme fonctionnelle d'un modèle agrégé de risque routier*. Actes Inrets n° 90 du groupe de travail 2001 du Séminaire Modélisation du Trafic, Arcueil.

BERGEL R., DEPIRE A. (2004b) *Climate, road traffic and road risk – an aggregate approach*. Proceedings of the WCTR'04, Istanbul.

BERGEL R., GIRARD B. (1993). “Un modèle de suivi des indicateurs journaliers d'insécurité routière” *Actes du Séminaire 1991-1992 sur la modélisation de l'insécurité routière*, Arcueil, INRETS., Tome 1, 102p., (Collection transports & communications N°47).

BERGEL R., GIRARD B., LASSARRE S., LE BRETON P. (1994) Un modèle de désaisonnalisation des indicateurs journaliers d'insécurité routière. *Actes du Séminaire 1993 de l'Inrets sur la Modélisation de l'insécurité routière*. Arcueil, INRETS, Tome 2, 152p., (Collection transports & communications N°48).

BERGEL R., GIRARD B., LASSARRE S., LE BRETON P. (1995a) *Un modèle de désaisonnalisation des indicateurs d'insécurité routière*, Rapport du groupe de travail OEST/CEMS/INRETS/SETRA, Paris, OEST.

BERGEL R., GIRARD B., LASSARRE S., LE BRETON P. (1995b) “A model of seasonally corrected road safety indicators”, Proceedings of the Conference Road Safety and Strategic Highway Research Program (SHRP) and Traffic safety on two Continents, Prague.

BERGEL R., GIRARD B., LASSARRE S., LE BRETON P. (1995c) “Un modèle de désaisonnalisation des indicateurs journaliers d'insécurité routière”, Notes de synthèse de l'OEST n° 93 – 09/1995

BERGEL R., MIGNOT V. (1992) “Une modélisation à court terme de la circulation routière”. Notes de synthèse de l'OEST n° 59 – 04/1992

BERGEL R., MUTTER A. (2000) “*Modélisation trimestrielle des trafics de voyageurs et de marchandises*”. Rapport de convention SES/INRETS n° 99/5814, Arcueil.

BIJLEVELD F.D., COMMANDEUR J.J.F., GOULD P.G., KOOPMAN S.J. (2008) “Model-based measurement of latent risk models in time series with applications”. *Journal of the Royal Statistical Society Series A* 171(1), 265-277.

BLARDONE J.D. (2007) “Modélisation des flux longue distance, le modèle Modev rénové”. Notes de synthèse du SESP n° 164 – 01/2007

BORGARD F., GUEGAN D. (1996) “Etudes de séries chronologiques linéaires à temps discret, Comparaison de logiciels”. *Rev. Statistique Appliquée*, XLIV(4), 59-80

BOURBONNAIS R., GRANGER T. (2002) Note interne de l'Université Paris-Dauphine.
www.securite-sanitaire.org/route/noteBourbonnais.htm

BOX G.E.P., COX D.R. (1964) “An analysis of transformations”. *Journal of the Royal Statistical Society*, B(2):211-243.

BOX G.E.P., JENKINS G.M. (1976) *Time series analysis. Forecasting and control*. Holden Day.

BOX G.E.P., TIAO G.C. (1975) “Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*. 70, 349, pp.70-79.

des transports de marchandises, SAE, Ministère des Transports.

BROCKWELL P.J., DAVIS R.A. (1991) *Time series : theory and methods*, second edition, Springer Verlag.

BROCKWELL P.J., DAVIS R.A.. (2002) *Introduction to time series and forecasting*, second edition, Springer Verlag.

CHATARD D., REYNAUD C., SALINI P. (1980) PRETRAM , Modèle annuel de prévision du trafic de marchandises, SAE, Ministère des Transports.

CHATARD D., SALINI P., HARCOUET C. (1981) PRETRAM , Reprise et actualisation des équations du modèle de prévision du trafic de marchandises, SAE, Ministère des Transports.

COMMANDEUR J., JOOPMAN S.J. (2007) *An introduction to state space time series analysis*, Oxford University Press.

COMMANDEUR J., BJLEVELD F., BERGEL R. (2007) *A multivariate time series analysis applied to SafetyNet data*. SafetyNet Deliverable D7.7 <http://www.erso.eu>

COST 329 (2004) *Models for traffic and safety development and interventions*. European Commission, Directorate general for Transport, Brussels , final report of the Action.

DAGUM E.B. (1975), “Seasonal factor forecasts from ARIMA models”, *Bulletin of the International statistical Institute* , N°46, pp203-16.

DAGUM E.B., QUENNEVILLE B. (1988), “Deterministic and stochastic models for the estimation of trading-day variation”, *Fourth Annual Research Conference*, Washington, D.C., Bureau of the Census, US Department of Commerce, pp569-90.

DARGAY J., HANLY M. (2002) “The demand for local bus services in England” *Journal of Transport Economics and Policy*, Volume 36, Number 1, pp.73-01(19)

DESAI U.B., PAL D., KIRKPATRICK R. (1985) “A realization Approach to stochastic model reduction” *Int. J. Control*, Vol. 42, n°4, 821-838.

DOBSON, A.J. (1990) “An introduction to Generalized Linear Models ”. Second Edition, Chapman and Hall, London.

DUPONT, E. ET MARTENSEN, H. (Eds.) (2007) *Multilevel modelling and time series analysis in traffic safety research - Methodology and Manual*. Deliverables D7.4 and D7.5 of the EU FP6 project SafetyNet <http://www.erso.eu>

DURBIN J, KOOPMAN S.J. (2001) *Time Series Analysis by State Space Methods*, Oxford University Press

ENGLE R.F., GRANGER C.W.J. (1987) Co-integration and error correction: Representation, estimation and testing, *Econometrica*, vol 55, 251-276.

FRIDSTROM L. et al. (1995) “Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts”, *Accident Analysis & Prevention*, Vol. 27 n°1 p.1-20.

GAUDRY M. (1984) *DRAG, un modèle de la Demande Routière, des Accidents et de leur Gravité, appliqué au Québec de 1956 à 1982*, Publication 359, Centre de Recherche sur les Transports, Université de Montréal.

GAUDRY M., LASSARRE S. (Eds.) (2000) *Structural Road Accident Models - The International DRAG Family*, Pergamon.

GUILAIN F. (1993) “Une approche qualitative de la météorologie en accidentologie routière”, *Actes du Séminaire sur la modélisation de l'insécurité routière*, Arcueil, I.N.R.E.T.S., Tome 1, pp. 59-66., (Collection transports & communications N°47).

GOODWIN, P. (1992) A review of new demand elasticities with special reference to short and long run effects of prices changes, *Journal of transport Economics and Policy*, 26, pp. 155-163.

GOODWIN P., DARGAY J., HANLY M. (2004) “Elasticities of road traffic and fuel consumption with respect to price and income : a review”, *Transport Reviews*, Vol. 24, n° 3, 275-292..

GOURIEROUX C., MONFORT A. (1990) *Séries temporelles et modèles dynamiques*, Economica.

HADDAK M., BERGEL R., LEJEUNE P., TRENY V., ARMOGUM J. (2005) *Risk Exposure Data in France. CETE-INRETS. Working Paper. Safetynet Task 2.1.*

HAKIM S., HAKKERT S., HOCHERMANN I., SHEFERT D. (1991) "A critical Review of macro models for road accidents". *Accident Analysis & Prevention*. Vol. 23, N° 5, pp. 379-400.

HARVEY A.C. (1989) *Forecasting structural time series and the Kalman filter*. Cambridge University Press, Cambridge.

HARVEY A.C., DURBIN J. (1986) "The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling", *J. R. Statist. Soc.*, Vol. 3 n°149 pp.187-227.

HILLMER S.C., TIAO G.C. (1982) "An ARIMA-model-based approach to seasonal adjustment", *J.A.S.A.*, Vol. 77, n° 377 pp. 63-70.

JAEGER L. (1998) *L'évaluation du risque dans le système des transports routiers par le développement du modèle TAG*. Thèse de doctorat ès sciences économiques, Université Louis Pasteur, Strasbourg.

JAEGER L., LASSARE S. (1999) "TAG, un modèle économétrique pour un suivi de l'insécurité routière" *RTS*, 65. pp.3-20

KHASNABIS S., LYOO S.H. (1989) Use of time series analysis to forecast Truck accidents *Transportation Research Record*, n° 1249 pp. 30-36, 1989.

LADIRAY D., QUENNEVILLE B. (2001) *Seasonal adjustment with the X11 method* Springer Verlag

LASSARRE S. (1994) *Cadrage méthodologique d'une modélisation pour un suivi de l'insécurité routière*, Synthèse INRETS, n°26, Arcueil.

LASSARRE S. (1996) Quelles tendances dans l'évolution du nombre de tués pour quatre pays: Finlande, France, Grande Bretagne et Pays-Bas ? Document COST 329, Inrets, Arcueil.

LASSARRE S. (2001) “ Analysis of progress in road safety in ten european countries ”. *Accident Analysis & Prevention*. Vol. 33, pp. 743-751

LE BRETON P., VERVIALLE F. “Un modèle de suivi de la sécurité routière”. Revue RTS. Arcueil: INRETS, 1990, n°27, pp. 33-42.

LENORMAND A. (2002) *Prévisions dans les modèles cointégrés avec rupture : application à la demande de transports terrestres de marchandises et de voyageurs*. Thèse de doctorat de l'Université Paris-1 Panthéon-Sorbonne.

MALAMOUD G. (1978) “Analyse de séries temporelles et prévisions : la méthode de Box et Jenkins ” *Annales des Ponts et Chaussées* – 3^{ème} trim. 1978

MCCULLAGH, P. AND NELDER, J.A. (1989) “Generalized Linear Models”. Second Edition. Chapman Hall, New York

MEYER K. (1998) *La concurrence rail-route : analyse économétrique des trafics de marchandises et des perspectives du transport combiné*. Thèse de doctorat de l'Université Paris-1 Panthéon-Sorbonne

MIGNOT V. (1991) « Modélisation de séries mensuelles de circulation routière – Modèles ARMAX avec variables météorologiques et calendaires ». Rapport de stage Jeunes-Etudes, OEST, Paris.

MUZEREAU E. (2004) *Analyse et prévision en présence de rupture du degré d'intégration : application aux trafics de la RATP*. Thèse de Doctorat de l'Université Paris-1 Panthéon-Sorbonne.

MORELLET O. (2007) *Les ménages et le transport dans le modèle M.A.T.I.S.S.E - Analyse rétrospective et prospective de l'équipement automobile et de la mobilité dans un pays tel que la France*. Rapport INRETS n° 273, Arcueil.

NALLET S. (2000) “Modélisation de séries journalières de sécurité routière ” Rapport de stage ISUP, Arcueil.

OEST (1987 à 1992) Notes de synthèse. Ministère de l'Équipement, La Défense

ONISR. *La sécurité routière en France : bilan de l'année* (2006) Ministère de l'Équipement, La Défense.

OPPE S. (1993) “The development of traffic and traffic safety in six developed countries”, *Accident Analysis & Prevention*, Vol. 23 n°5 p.401-412.

OPPE, S., KOORNSTRA M.J. (1990) A mathematical model for related long term developments of road traffic and safety. In: Koshi, M. (Ed.), *Transportation and Traffic Theory*. Elsevier, New York, pp. 113-133.

PELZMAN, S. (1975) “The effects of automobile safety regulation”. *Journal of Political Economy*, 83(4), 677-725.

PINON B , BERGEL R., CREBASSA P. (1995) “*Constitution d'un outil de suivi et de prévision à court terme du trafic aérien domestique*”. OEST, Ministère de l'Équipement, La Défense.

RUCKEBUSCH G. (1978) Représentations markoviennes de processus gaussiens stationnaires et applications statistiques. *Mathematics*, 636 : 115-139.

SAEP-OEST (1984 à 1992) “Les projections semestrielles de transport” Responsable de la publication et modélisation : Ruth Bergel. Ministère de l'Équipement, La Défense.

SESP (2008a) Notes de conjoncture du SESP, Collection Infos rapides,
<http://www.statistiques.developpement-durable.gouv.fr>

SESP (2008b) Dossier des notes de synthèse. OEST - SES – SESP. Modèles économiques dans les domaines des transports et du logement 1988-2008
<http://www.cdat.developpement-durable.gouv.fr>

SMEED, R.J. (1949) Some statistical aspects of road safety research. *Journal of the Royal Statistical Society A* 112(1), pp.1-24

SCOTT P. P. (1986) "Modelling Time-Series of British Road Accident Data", *Accident Analysis & Prevention*, Vol. 18 n°2 pp.109-117.

STIPDONK H. (Ed.) (en cours de publication 2008). Deliverable D7.10 of the EU FP6 project SafetyNet <http://www.erso.eu>

TAROUX J.P., BUCHMULLER G. (1980), PRETRAP, Modèle de prévision des transports de personnes, SAE, Ministère des Transports.

TAROUX J.P (1981) PRETRAP, Modèle macro-économique de prévision des Transports de personnes, Note de synthèse, SAE, Ministère des transports.

VIZATELLE C. (1990) Modélisation de séries mensuelles de sécurité routière – Modèles ARMAX avec variables météorologiques et calendaires. Rapport de stage Jeunes-Etudes, OEST, Paris.

WAGENAAR, A.C., MAYBEE R.G. (1986) "The legal minimum drinking age in Texas: effects of increase from 18 to 19", *Journal of safety research*, Vol. 17 pp.165-178.

WILDE T.J. (1982) "The theory of risk-homeostasis: implications for safety and health", *Accident Analysis and Prevention*, Vol. 2 pp..

WIRKOWSKY J.J., HECKARD R.E. (1977) "The use of time series analysis and intervention analysis to assess the effects of external factors on traffic indices: a case study of the effects of speed limit reduction and energy crises in the state of Texas", *Accident Analysis & Prevention*, Vol. 9 n°1 pp.229-247.

YANNIS G et al. (2005) *State of the Art Report on Risk and Exposure Data*. SafetyNet Deliverable 2.1 <http://www.erso.eu>

LA PRISE EN COMPTE DE VARIABLES EXPLICATIVES DANS LES MODELES DE SERIES TEMPORELLES : APPLICATIONS A LA DEMANDE DE TRANSPORT ET AU RISQUE ROUTIER

Résumé

L'objet de la thèse est d'exposer une démarche méthodologique qui vise à prendre en compte, dans les modèles de séries temporelles, des effets exogènes mesurés à l'aide de variables additionnelles, et de l'illustrer par un certain nombre d'applications au secteur des transports. Dans ces applications, le pas de temps est le jour, le mois, le trimestre voire le semestre : il s'agit de prendre en compte des effets exogènes, de nature transitoire ou de nature durable, qui se manifestent dans le court terme.

La première partie de la thèse traite de la modélisation des séries temporelles. Nous situons le cadre formel des modèles auxquels nous nous intéressons, nous exposons la démarche suivie dans le cadre des modèles ARMA avec variables explicatives, puis dans le cadre des modèles markoviens avec variables explicatives en y détaillant le cas particulier des modèles structurels.

Les deuxième et troisième parties de la thèse regroupent deux ensembles d'applications. Le premier porte sur des données de trafics, de voyageurs et de marchandises, agrégées par mode de transport ou par grande catégorie de réseau, et le second sur des données d'accidents corporels et de victimes de la circulation routière, agrégées par grande catégorie de réseau routier. La période couverte la plus large est 1970-2000.

La plupart des applications intègre la prise en compte des effets transitoires, de nature climatique et calendaire, sur la demande de transport et sur le risque routier, et nous donnons dans la thèse les premiers résultats détaillés démontrant pour la France la significativité du facteur climatique sur le bilan routier national, mesuré en nombres d'accidents corporels et de tués.

Mots-clefs

Modèles de séries temporelles. ARMA, espace état, variables explicatives, climat, calendrier, demande de transport, risque routier.

THE USE OF EXPLANATORY VARIABLES IN TIME SERIES MODELLING: APPLICATIONS TO TRANSPORT DEMAND AND ROAD RISK

Abstract

The aim of the thesis is to set out a methodology that includes in time-series modelling exogenous effects measured by additional variables. This methodology is illustrated by a number of applications relating to transport. In these applications, time is measured in days, months, quarters and semesters (half years). We aim to take account of exogenous effects which are either transitory or durable lasting and which manifest themselves in the short term

The first part of the thesis deals with time-series modelling. We provide a typology of time-series models and place our approach within it. We describe the approach used in ARMA modelling with explanatory variables and then in state space modelling with explanatory variables, paying special attention to structural time-series modelling.

The second and third parts bring together two groups of applications. The first group considers traffic datasets, for passengers and for freight, aggregated by mode and by main network type. The second group considers numbers of road injury accidents and casualties, aggregated by main network type. The largest period covered is 1970-2000.

Most of the applications address the transitory effects on transport demand and road risk of weather and calendar factors. We provide the first detailed results that demonstrate the significance of weather factor on road safety in France, measured by numbers of injury accidents and fatalities.

Key Words

Time series modelling, ARMA, state space, explanatory variables, weather, calendar, transport demand, road risk.