

## Annexe I - Origines de la stabilité

Les pistes des origines de la stabilité des budgets-temps de transport, évoquées dans cette partie, sont apportées par différents champs disciplinaires. Selon les disciplines, les propositions sont le plus souvent difficilement testables. De façon générale, elles s'organisent autour des trois axes : les solutions biologiques, les solutions économiques et les solutions sociologiques (Baveling et al., 2001).

Les premières rassemblent les explications produites par les disciplines telles que l'éthologie, la sociobiologie, la psychologie. Elles concernent essentiellement les régularités comportementales propres à la nature humaine et peuvent être assimilées à une détermination génétique du comportement. Les solutions économiques font référence aux disciplines qui dressent un cadre d'hypothèses relatives à la nature de l'activité humaine, afin d'expliquer un comportement. La plupart suppose que l'individu adopte un comportement rationnel. Enfin, les solutions sociologiques regroupent les explications sociologiques, anthropologiques et historiques et expliquent le comportement par le contexte de l'activité humaine.

### I. Les solutions biologiques

Ce groupe d'explications de la stabilité est composé de solutions de type biologique, voire génétique (Baveling et al., 2001), anthropologique (Marchetti, 1994) et bio-psychologique (Michon, 1980).

Les anthropologues ont avancé que la culture est un déterminant principal du comportement humain. Ainsi, l'observation de « loi » universelle est d'autant plus rare que les cultures diverses. Mais, pour certains, l'influence exercée par notre culture sur le comportement n'est pas suffisante pour nous éloigner de nos comportements naturels. Ce rejet de l'indépendance du comportement par rapport à notre constitution biologique est ainsi partagée par de nombreuses branches scientifiques (sociobiologie, anthropologie Darwiniste, etc.). Les sociobiologistes et certains psychologues ajoutent aux côtés des déterminants culturels des déterminants biologiques (Wilson, 1975). Une part génétique peut donc être à l'origine de la régularité du comportement. L'étude d'un comportement se concentre donc sur les questions :

- de la part héréditaire du comportement. Jusqu'à quel point le comportement est-il inscrit dans nos gènes ?
- des objectifs et des fonctions du comportement. Jusqu'à quel point nos comportements sont-ils liés à d'anciens objectifs de conservation, de procréation, etc. ?
- des mécanismes du comportement. Quels sont les stimuli qui font naître et qui façonnent un comportement ?
- de développement du comportement durant la vie de l'individu. Le comportement est, pour partie, acquis au travers d'un processus d'apprentissage de génération en génération.

#### 1. La survie par la mobilité

Pour un certain nombre d'animaux, les « temps de déplacement » observés sont relativement stables. Pour répondre à leurs besoins vitaux, ils doivent chasser, cueillir et donc se déplacer. La constance du temps alloué à certaines activités n'est pas rare chez les animaux. Ainsi, les chiens sauvages ou certains faucons passent un temps constant à chasser. Et les animaux domestiques, comme les canaris ou les hamsters, passent une durée stable à courir ou voler durant la journée.

Les raisons de ces stabilités sont étudiées par les biologistes. Tout d'abord, il est avancé que les animaux et les hommes ont des stocks limités d'énergie, de nourriture et de temps. Une bonne gestion de ces ressources est vitale pour un animal sauvage. Il est risqué de dépenser toute son énergie en peu de temps. Les animaux doivent gérer leurs efforts à la fois pour maintenir leur niveau de ressources vitales, mais aussi pour maintenir leur forme physique. C'est ce qui explique pourquoi les temps alloués à certaines activités, dont la mobilité, ont des bornes supérieures et inférieures. Ensuite, la gestion du risque semble faire intervenir le même type d'arbitrage, entre d'une part le besoin de protection et d'autre part, le plaisir de l'excitation naissant avec la prise de risque.

Ces deux explications reposent sur le concept de la « sélection naturelle » provenant de la théorie de l'évolution. Comme la sélection naturelle est un processus s'étalant sur plusieurs générations, il est supposé que la nature humaine s'est développée durant l'ère Paléolithique. Comparée à cette longue période, notre histoire récente n'est pas assez longue pour que notre constitution biologique ait pu être modifiée. Ainsi, nos comportements sont fondamentalement très proches de ceux de nos ancêtres (Burnham et Phelan, 2000).

## 2. L'instinct de mobilité

Le besoin de ressources explique certainement une part de notre besoin de mobilité. Mais l'activité de recherche explique elle aussi notre mobilité actuelle. Notre instinct s'est développé à une période où les ressources étaient rares. Mais maintenant, dans les pays développés, les ressources sont abondantes et facilement accessibles. C'est simplement la recherche de ressources qui nous rend heureux et non leur possession (Burnham et Phelan, 2000). La satisfaction de ce comportement inscrit dans nos gènes persiste et motive notre mobilité. Notamment, les biologistes parviennent à identifier les gènes responsables de notre besoin de mobilité et de découverte (LeMoal et Simon, 1991). Il semble que certaines dopamines intensifient notre besoin de mobilité (Fink et Smith, 1980 ; Hamer et Copeland, 1998).

## 3. Bio-psychologie : la recherche d'un équilibre plutôt que d'une satisfaction

L'étude de la motivation du comportement par la bio-psychologie apporte un double éclairage. En effet, tout d'abord, la biologie avance que l'homme partage avec les grands animaux le besoin d'explorer son environnement et de disposer de ressources suffisantes et variées. Ensuite, la psychologie constitue un cadre explicatif des motivations du comportement. Ainsi, Michon (1980) suppose que l'homme s'adapte à son environnement et qu'il cherche à préserver un équilibre face à des situations variables. Cette hypothèse de représentation du comportement a notamment été utilisée par Sciotovsky (1976), qui supposait comme première motivation du comportement humain, le maintien d'un niveau optimal d'éveil ou d'activité, plutôt que la recherche de satisfaction. Ainsi, le comportement est gouverné par les expériences passées et notamment celles apportant du plaisir en déviant du niveau optimal et habituel d'éveil. Ainsi, l'homme doit répondre à un besoin de mobilité et d'exploration, mais sans trop s'écarter de son habitude : un budget-temps de transport d'une heure. Le retour à un niveau habituel de mobilité est alors motivé par le stress. Ainsi, les comportements se modifient dans les grandes villes, où les distances à parcourir sont importantes et peuvent amener à dépasser le budget-temps de transport. Des comportements de substitution sont développés, de type récréatif ou résidentiel. Afin de réduire le stress du long trajet, le temps du transport est agrémenté d'autres activités distrayantes ou similaires à celles du domicile. La technologie embarquée dans les modes de transport et les

comportements « non-transport » observés dans les transports en commun en sont des indicateurs. Dans les cas extrêmes où le temps de transport est trop éloigné d'une heure, et où les comportements de substitutions ne peuvent être développés, les localisations résidentielles, les emplois ou les modes de transport sont modifiés.

La régularité comportementale imposée par la nature humaine est la base du raisonnement de Marchetti (1994). Dans son exposé, l'homme est un « animal de territoire », qui est motivé par un instinct d'extension de sa zone d'activité, de laquelle dépend son niveau de ressources. Il est aussi caractérisé par un « instinct de caverne » qui le pousse à retourner dans sa grotte à la fin de la journée. Pour Marchetti, l'hypothèse de Zahavi montre « l'unité quintessentielle de l'instinct de mobilité autour du monde »<sup>91</sup>, qui est validée au-delà des cultures, des races et des religions. La stabilité du budget-temps de transport résulte des instincts hérités de nos ancêtres, qui nous poussent à explorer un territoire et à retourner dans nos grottes le soir. L'équilibre obtenu détermine un temps « d'exposition » qui est proche d'une heure.

Sous cette hypothèse, Marchetti justifie le raisonnement selon lequel la structure et la croissance des villes sont expliquées par les vitesses disponibles et le budget-temps de transport. La suprématie des instincts biologiques permet alors de définir la zone d'activité humaine en fonction du temps d'exposition décelé par Zahavi et selon les vitesses de déplacement. Ainsi, à la vitesse de la marche à pied, de 5 km/h pendant une heure et en « rentrant à la grotte » pour la nuit, le territoire qui peut être parcouru a un rayon de 2,5 km, soit une surface de 20 km<sup>2</sup>. Ces dimensions spatiales sont retrouvées pour les villages où la marche à pied domine et aussi, jusqu'au 19<sup>ème</sup> s. pour les villes fortifiées, dont les murailles n'avaient pas de diamètre supérieur à 5 km. L'avènement successif des modes de transport plus rapides pourrait alors expliquer l'extension spatiale des villes. Les vitesses supérieures repoussant alors les frontières des zones d'activités humaines.

---

<sup>91</sup> Marchetti (1994), p. 1-2.

## II. Les solutions économiques

L'essentiel des explications économiques relatives aux choix de transport considèrent l'homme non pas comme une entité biologique, mais comme un individu optimisant son utilité sous un certain nombre de contraintes. Les modèles économiques de transport supposent que le transport est une demande dérivée, qui acquiert la valeur de l'activité à destination. De ce fait, le temps consacré au transport est ressenti comme un coût que l'individu cherche à minimiser. L'individu recherche donc une gestion optimale des ressources engagées dans chacun de ces choix, principalement le temps et l'argent pour les choix de transport.

La première explication économique de la stabilité des dépenses de transport a été énoncée par Hupkes (1982). Les caractéristiques du choix de transport sont considérées comme le résultat d'un processus d'optimisation. L'homme est donc considéré rationnel et cherche à maximiser l'utilité de sa mobilité, qui peut être décomposée en deux termes. Une utilité *dérivée* qui tient compte de la nature dérivée du transport. Elle correspond donc à l'utilité de l'activité à destination. Elle est tout d'abord croissante, car la zone d'activités accessibles s'étend avec le temps consacré au transport. Le nombre d'opportunités socio-économiques accessibles augmente. Puis, elle est décroissante car le temps de transport est de plus en plus en concurrence avec le temps pour les autres d'activités. A cette utilité dérivée, s'ajoute une utilité *intrinsèque* correspondant au plaisir résultant du transport pour lui-même. Le mouvement offre certains plaisirs, tels que le changement de cadre, la sensation de vitesse, le sentiment de liberté, l'excitation de la conduite d'un véhicule, ou tout simplement le plaisir du mouvement (Mokhtarian et Salomon, 2001). Cette utilité intrinsèque est croissante, puis rapidement décroissante, voire négative lorsque le transport, le mouvement, devient un effort non désiré.

Ce raisonnement économique montre donc l'existence d'un niveau optimum de transport. Mais, il ne suppose pas la constance d'un temps de déplacement. Cette dernière proviendrait de la répétition d'un choix d'allocation dans un environnement stable. Au contraire, Hupkes proposait de rechercher les facteurs susceptibles d'affecter l'une ou l'autre des deux utilités, donc de rechercher les éléments intervenant dans la formation de l'univers de choix de l'individu.

Reprenant l'hypothèse de l'utilité intrinsèque du transport, Mokhtarian et Salomon (2001), décomposent l'utilité du transport en trois éléments :

- l'utilité de la destination. Elle est équivalente à l'utilité dérivée de Hupkes,
- l'utilité de l'activité menée pendant le déplacement,
- l'utilité produite par le déplacement. Elle correspond au simple plaisir du mouvement (plaisir de la vitesses, sensation de liberté, etc.).

Ainsi, ils définissent plusieurs sources positives d'utilité du transport dont notamment une partie qui ne correspond pas à la nature purement dérivée de la demande de mobilité. Mokhtarian et Salomon supposent alors que les individus ne minimisent pas uniformément leurs dépenses temporelles de transport. Mais qu'il existe un niveau désiré de mobilité inobservé, qui varie entre les individus et pour un même individu dans le temps. Si le budget-temps de transport réalisé excède ce niveau désiré, l'individu voudra diminuer sa dépense temporelle. Si le budget-temps de transport réalisé est inférieur au niveau désiré, l'individu voudra augmenter son temps de transport. Enfin, le niveau de ce budget-temps de transport désiré varie avec les caractéristiques personnelles, les attributs démographiques, les modes de transport utilisés et les motifs des déplacements.

Sous cette hypothèse, les stabilités observées peuvent provenir du fait que les fluctuations autour du niveau idéal de mobilité se compensent. De plus, les variations des budgets-temps de transport observées au niveau micro peuvent s'expliquer par le fait que le niveau désiré peut être une fonction des caractéristiques individuelles, et qu'il n'est pas toujours défini selon les contraintes pesant sur la mobilité.

### III. Les solutions sociologiques

L'étude sociologique se concentre sur le contexte des choix effectués par les individus. Pour identifier ce qui conduit les individus à se comporter d'une façon ou d'une autre, la sociologie recherche les motivations et le contexte individuels. Ces motivations diffèrent dans le temps, dans l'espace et d'une situation à l'autre. Ainsi, il n'existe pas ou très peu de règles pouvant être appliquées de façon universelle. Un schéma explicatif de la constance pour une population donnée ne pourra être soutenu face à la diversité des contextes locaux et historiques.

De plus, la stabilité du temps de transport fait l'hypothèse que le temps est mesurable, divisible, calculable et universel. Or, les psychologues et sociologues du temps ont montré que les perceptions du temps diffèrent entre les pays et les populations. Dans les sociétés modernes, le temps est de plus en plus souvent multiple. Ainsi, les comparaisons des perceptions du temps entre des cultures différentes ou pour un même individu à des périodes différentes sont extrêmement difficiles.

Le temps de transport en tant que choix individuel trouve relativement peu d'explication dans la sociologie. Ses références au contexte de la constitution d'un comportement et notamment aux motivations et aux perceptions qui caractérisent l'individu, rendent problématique la généralisation d'un schéma explicatif à l'ensemble d'une vaste population caractérisée par ailleurs par de fortes différences culturelles.

En définitive, seul le caractère habituel d'un comportement sociologique pourrait expliquer cette stabilité. La routine peut s'inscrire sur la gestion du temps de transport, mais aussi sur l'ensemble des temps. Hägerstrand (1973) suggérait, pour une personne active, le découpage des temps suivant : 9 heures de sommeil et de soins de soi, 1 heure de sustentation, 8 heures de travail et 2 heures pour les achats et les activités de services. Les 4 heures restantes peuvent alors être consacrées au transport et aux loisirs. Sous cet ensemble de contraintes temporelles, il n'est pas surprenant que les temps de transport moyens se concentrent autour d'une moyenne d'un peu plus d'une heure. Le caractère routinier de ce découpage temporel peut être utilisé pour énoncer l'emploi du temps représentatif des populations.

Ce n'est alors pas seulement les origines du temps de transport qui sont à rechercher, mais les origines de l'ensemble des durées des activités. Il est bien évident que pour ce problème plus général, la diversité des contextes, des perceptions, des motivations, etc. empêchent de dresser un cadre universel explicatif de la stabilité d'un tel emploi du temps.

## Conclusion

L'ensemble de ces explications fait référence au comportement individuel avec l'objectif d'expliquer la stabilité de la moyenne. Le passage de l'unité individuelle à la moyenne est le point critique de l'explication de la stabilité.

L'explication biologique avance que l'humain est programmé pour se déplacer une certaine partie de sa journée. Mais pourquoi observe-t-on alors de si grandes différences selon les modes de transport, selon les jours, les saisons et les cycles de vies, etc. ? La réponse économique suppose que l'utilité nette du transport est croissante jusqu'à un certain temps de déplacement, puis décroissante car le coût généralisé est supérieur à l'utilité du transport. Mais comment ce niveau optimal peut-il être identique pour tous les individus ? Les accessibilités sont totalement différentes suivant le lieu de résidence, le moyen de transport, etc. (Hupkes, 1982). De plus, la valorisation de ces accessibilités diffère selon les individus et selon plusieurs échelles temporelles pour un même individu (Höjer et Mattson, 2000). De nombreux facteurs peuvent affecter les utilités intrinsèque et dérivée. Parmi ceux-ci des changements sociétaux de grande ampleur tels que les réductions de temps de travail ou une substitution majeure entre les nouvelles technologies de l'information et de la communication et les transports pourraient modifier les formes de ces utilités et donc le temps de transport optimal. De même, les explications sociologiques peuvent difficilement être supposées stables entre individus. L'observation de Hägerstrand (1973) correspond au découpage temporel d'un homme de l'époque contemporaine. Mais, elle est difficilement transférable à toutes les populations pour lesquelles la stabilité du budget-temps de transport est observée. De grandes variations peuvent être observées autour de cet emploi du temps moyen au niveau désagrégé pour les populations inactives, les populations rurales, etc. Et au niveau agrégé, les structures de ces sous-populations ne sont pas identiques et les organisations temporelles diffèrent, par exemple les japonais travaillent environ 25% de temps en plus que les européens.

Loin de supposer une identité de comportements face au choix du temps de transport, la conjecture de Zahavi mentionne l'existence de variations autour de la moyenne. La composition de plusieurs situations produit la stabilité observée de la moyenne. Il faut donc rechercher les origines de la stabilité des temps de transport, tout d'abord, dans les mécanismes de la constitution de la moyenne par agrégation des comportements individuels.

Ensuite, la stabilité trouvera une explication par l'identification et l'explication d'un comportement moyen conduisant à une dépense de transport stable pour les différents groupes d'individus.

Enfin, l'étude des variations autour de la moyenne révélera très certainement de nombreuses informations tant sur les mécanismes intervenant dans la constitution de la moyenne (compensation, effet de structure, etc.), que sur les comportements individuels.

## Annexe II - L'enquête ménages de Lyon 1995

### Présentation de l'enquête ménages de Lyon 1995

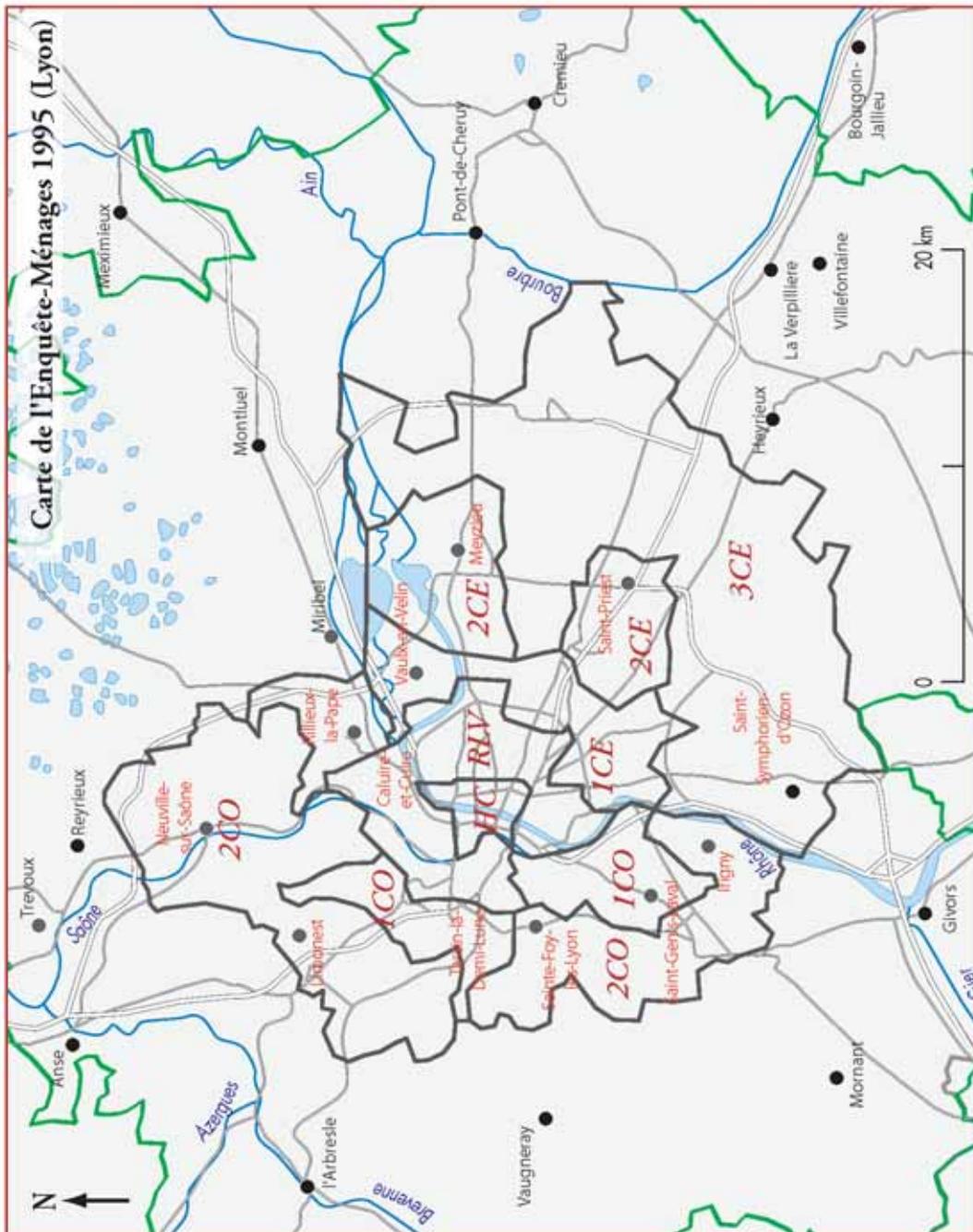
L'enquête ménages déplacements (CERTU, 1998) constitue la principale source de données française en milieu urbain sur les déplacements réalisés au sein d'une agglomération. Dans l'objectif de produire une information aidant à la réflexion dans le domaine de la planification des déplacements urbains, l'enquête ménages fournit une connaissance fine des déplacements. Pour l'ensemble des agglomérations de province, une méthodologie standard est appliquée. La population cible est constituée de l'ensemble des déplacements réalisés par les résidents du périmètre d'étude. L'échantillon est constitué par tirage aléatoire de ménages, après stratification géographique du périmètre d'enquête, au sein de la base ménages du recensement de 1990, mis à jour à partir du fichier des constructions neuves.

Tous les individus du ménage de 5 ans et plus sont enquêtés le même jour sur l'ensemble de leurs déplacements réalisés la veille du jour d'enquête. L'enquête ne porte que sur des jours de semaine. Les déplacements recensés sont ceux effectués sur la voie publique, à l'intérieur du périmètre d'enquête entre 4 heures du matin la veille et 4 heures du matin le jour même. Un déplacement est défini comme : « un aller simple effectué d'un point (zone origine) à un autre point (zone destination), pour un motif unique, selon une heure de départ et une heure d'arrivée et effectué avec un ou plusieurs modes de transport » (Cete de Lyon et al., 1995). 13997 individus sont enquêtés, issus de 6001 ménages, pour 53213 déplacements.

### Les limites des enquêtes ménages françaises

Malgré la similitude de la méthodologie dans le temps, le dispositif des enquêtes ménages admet certaines critiques et limites (Bonnell, 2004). Tout d'abord, l'enquête ménages se concentre sur la mobilité des résidents du périmètre d'enquête. De fait, une partie des déplacements au sein de ce périmètre sont exclus, lorsqu'ils sont réalisés par des non-résidents de la zone. Ainsi, les déplacements de transit et les déplacements professionnels ne sont pas pris en compte. Certaines évaluations (Nicolas et al., 2001 ; Orfeuil, 2002) ont notamment montré l'importance de ces omissions (plus d'un tiers des véhicules kilomètres parcourus un jour moyen de semaine). Ensuite, le périmètre des enquêtes ménages reste

encore parfois trop restreint par rapport au bassin d'attraction de l'agglomération. L'étalement urbain, le desserrement des activités et le développement des franges des zones urbaines élargissent peu à peu cette zone, qui nécessite une adaptation du périmètre d'enquête. Enfin, Bonnel (2004) rappelle que la taille d'échantillon est souvent faible pour permettre l'analyse désagrégée selon certaines dimensions. En conséquence pour constater des différences significatives entre segments de population une taille d'échantillon suffisante est nécessaire.



source : d'après l'Atlas du Grand Lyon, INSEE, Cartographie et Décision, 1994

## Annexe III - Les budgets-temps de transport de l'Enquête-ménages de Lyon 1995

### Statistiques descriptives des budgets-temps de transport corrigés

99 individus sont écartés en raison de durées de déplacement incorrectes.

Variable d'analyse : Budget-temps de transport						
Temps de transport erroné	Nb obs.	Nb	Moyenne	Écart Type	Minimum	Maximum
0	12364	12364	79.19	62.35	2.0	960.0
1	99	99	222.07	461.41	0	1675.0

1062 individus sont écartés en raison de déplacements dont l'origine et/ou la destination sont en dehors de l'aire urbaine définie par l'enquête.

Budget-temps de transport						
Hors Zone	Nb obs.	Nb	Moyenne	Écart Type	Minimum	Maximum
0	11401	11401	74.46	65.37	0	1555.0
1	1062	1062	143.33	128.81	10.0	1675.0

120 individus sont écartés en raison de déplacements qui ne sont associés à aucun trajet de retour à domicile.

Budget-temps de transport						
Un seul déplacement	Nb obs.	Nb	Moyenne	Écart Type	Minimum	Maximum
0	12343	12343	79.37	70.08	2.0	1675.0
1	120	120	178.71	278.77	0	1500.0

84 individus sont écartés en raison de budgets-temps de transport supérieurs à 6 h.

Budget-temps de transport						
BTT > 360	Nb obs.	Nb	Moyenne	Écart-type	Minimum	Maximum
0	12379	12379	76.55	52.15	0	360.0
1	84	84	637.40	364.28	365.0	1675.0

1534 individus sont immobiles et ne sont donc pas considérés.

<b>Budget-temps de transport</b>						
<b>Immobile</b>	<b>Nb obs.</b>	<b>Nb</b>	<b>Moyenne</b>	<b>Écart Type</b>	<b>Minimum</b>	<b>Maximum</b>
0	12463	12463	80.33	75.44	0	1675.0
1	1534	0	.	.	.	.

18 individus sont écartés en raison d'horaires d'activités érronés.

<b>Budget-temps de transport</b>						
<b>Horaire érroné</b>	<b>Nb obs.</b>	<b>Nb</b>	<b>Moyenne</b>	<b>Écart Type</b>	<b>Minimum</b>	<b>Maximum</b>
0	13979	12445	79.46	66.87	0	1500.0
1	18	18	680.94	718.38	10.0	1675.0

## Statistiques descriptives des budgets-temps de transport

### **Statistiques descriptives des budgets-temps de transport avant correction (en min)**

Moyenne	80.33	Médiane	65	Intervalle Interquartile	61
Ecart type	75.44	Mode	60	Intervalle des valeurs	1675
N	12463	Quantile 75% (Q3)	101	Quantile 25% (Q1)	40

### **Statistiques descriptives des budgets-temps de transport (en min)**

Moyenne	72.21	Médiane	60	Intervalle Interquartile	55
Ecart type	46.99	Mode	60	Intervalle des valeurs	353
N	11232	Quantile 75% (Q3)	95	Quantile 25% (Q1)	40

## Annexe IV - Définitions des variables

Variable	Définition
<i>Variables socio-économiques</i>	
Genre	Homme = 1 / femme = 0
Age	Age exact ou classes d'âges : de 0 à 19 ans ; de 20 à 49 ans ; plus de 50 ans.
Statut professionnel	Statuts codés de 1 à 9 : travail à temps plein, travail à temps partiel, étudiant, scolaire, chômeur, retraité, reste au foyer, formation ou stage, et autre
Actif	Actifs à temps complet ou partiel et les étudiants
Scolaire	Collégiens et lycéens
Inactif	Chômeurs et retraités
Femme au foyer	Femmes au foyer
Nombre de personnes totales du ménage	Comptabilise le nombre de membres du ménage
Nombre d'enfants	Le nombre d'enfants du ménage
Nombre d'enfants de moins de 5ans	Le nombre d'enfants du ménage de moins de 5ans
Nombre d'enfants de 6 à 17 ans	Le nombre d'enfants du ménage de 6 à 17 ans
Revenu du ménage	Le revenu mensuel du ménage est classé selon les paliers : 0 ; 2500 ; 5000 ; 7500 ; 10000 ; 12500 ; 15000 ; 20000 ; 30000 ; 50000 ; 50000 et plus (en F) Les ménages à hauts revenus sont définis comme supérieurs à 20000F. Les ménages à faibles revenus sont définis comme inférieurs à 5000F.
<i>Variables de transport</i>	
Titulaire du permis de conduire	Codée 1 si l'individu est titulaire du permis de conduire
Nombre de VP à libre disposition du ménage	Nombre de voitures particulières à libre disposition du ménage allant de 0 à 4 et plus.
Mode principalement utilisé	Codé de 1 à 7 : La marche à pieds, le vélo, les deux-roues, les transports en communs, la voiture particulière en tant que conducteur ou passager et la classe « autre ». Le mode est dit mode principal s'il représente la plus grande part du nombre de déplacements effectués dans la journée.

## Annexe IV - Définitions des variables

---

<b>Variable</b>	<b>Définition</b>
<i>Autres variables</i>	
Jour des déplacements	De 1 à 5 : du lundi au vendredi
Zone de domicile	Définie de 1 à 8 selon : hypercentre, Lyon-Villeurbanne, 1 <sup>ère</sup> couronne Est et Ouest, 2 <sup>ème</sup> couronne Est et Ouest, 3 <sup>ème</sup> couronne Est, et zone externe
<i>Variables du programme d'activités</i>	
Budget-temps de travail	Somme des durées quotidiennes de travail, à temps plein et à temps partiel et des durées à l'université.
Budget-temps de loisir	Somme des durées quotidiennes de activités : loisir sportif ; loisir culturel ; loisir associatif ; promenade ; restaurant ; visite d'amis
Budget-temps d'achat et service	Somme des durées quotidiennes de activités : achat quotidien ; achat hebdomadaire ; achat d'équipement, de vêtements et de loisir ; achat de santé ; démarches ; recherche d'emploi
Accompagnement	Codée 1 si l'individu réalise un déplacement d'accompagnement

## Annexe V - Composition de la base MCD

Population	
Surface area (ha)	
Characteristics of the metropolitan area	
Urban density	persons/ha
Job density	jobs/ha
Proportion of jobs in CBD	%
Metropolitan gross domestic product per capita	USD
Supply indicators	
Private Transport Infrastructure Indicators	
Length of road per 1000 people	m/1000 persons
Length of freeway per 1000 people	m/1000 persons
Length of road per urban hectare	m/ha
Length of freeway per urban hectare	m/ha
Parking spaces per 1000 CBD jobs	spaces/1000 jobs
Public Transport Infrastructure Indicators	
Total length of public transport lines per 1000 people	m/1000 persons
Total length of reserved public transport routes per 1000 people	m/1000 persons
* Bus reserved route length per 1000 people	m/1000 persons
* Minibus reserved route length per 1000 people	m/1000 persons
* Segregated tram network length per 1000 people	m/1000 persons
* Light rail network length per 1000 people	m/1000 persons
* Metro network length per 1000 people	m/1000 persons
* Suburban rail network length per 1000 people	m/1000 persons
* Heavy rail network length per 1000 people	m/1000 persons
Total length of reserved public transport routes per urban hectare	m/ha
* Bus reserved route length per urban hectare	m/ha
* Minibus reserved route length per urban hectare	m/ha
* Segregated tram network length per urban hectare	m/ha
* Light rail network length per urban hectare	m/ha
* Metro network length per urban hectare	m/ha
* Suburban rail network length per urban hectare	m/ha
* Heavy rail network length per urban hectare	m/ha
Intermodal Transport Infrastructure Indicators	
Number of park and ride facilities per kilometre of reserved public transport route	facilities/km
Number of park and ride spaces per kilometre of reserved public transport route	spaces/km
Number of park and ride facilities per 10,000 urban hectare	facilities/10000 ha
Car equivalents per number of park and ride spaces	units/space
Private transport supply [cars and motorcycles]	
Passenger cars per 1000 people	units/1000 people
Motor cycles per 1000 people	units/1000 people
Total private passenger vehicles per 1000 people	units/1000 people
Passenger car kilometres per car	km/unit
Motor cycle kilometres per motor cycle	km/unit
Total private passenger vehicle kilometres per vehicle	km/unit
Private collective transport supply [taxis and shared taxis]	
Taxis per million people	units/10 <sup>6</sup> persons
Shared taxis per million people	units/10 <sup>6</sup> persons
Taxi vehicle kilometres per capita	v.km/person

## Annexe V - Composition de la base

---

Shared taxi vehicle kilometres per capita	v.km/person
Traffic Intensity Indicators	
Passenger cars per kilometre of road	units/km
Motor cycles per kilometre of road	units/km
Total private passenger vehicles per kilometre of road	unit equivalents/km
Total single and collective private passenger vehicles per kilometre of road	unit equivalents/km
Passenger car kilometres per kilometre of road	v.km/km
Motor cycle kilometres per kilometre of road	v.km/km
Total private passenger vehicle kilometres per kilometre of road	v.km/km
Total private and collective passenger vehicle kilometres per kilometre of road	v.km/km
Passenger car kilometres per urban hectare	v.km/ha
Motor cycle kilometres per urban hectare	v.km/ha
Total private passenger vehicle kilometres per urban hectare	v.km/ha
Total private and collective passenger vehicle kilometres per urban hectare	v.km/ha
Average road network speed	km/h
Public Transport Supply and Service	
Total public transport vehicles per million people	units/10 <sup>6</sup> persons
* Buses per million people	units/10 <sup>6</sup> persons
* Minibuses per million people	units/10 <sup>6</sup> persons
* Tram units per million people	units/10 <sup>6</sup> persons
* Light rail units per million people	units/10 <sup>6</sup> persons
* Metro units per million people	units/10 <sup>6</sup> persons
* Suburban rail units per million people	units/10 <sup>6</sup> persons
* Heavy rail units per million people	units/10 <sup>6</sup> persons
Total public transport vehicle kilometres of service per capita	v.km/person
* Bus vehicle kilometres per capita	v.km/person
* Minibus vehicle kilometres per capita	v.km/person
* Tram wagon kilometres per capita	v.km/person
* Light rail wagon kilometres per capita	v.km/person
* Metro wagon kilometres per capita	v.km/person
* Suburban rail wagon kilometres per capita	v.km/person
* Heavy rail wagon kilometres per capita	v.km/person
Total public transport vehicle kilometres of service per urban hectare	v.km/ha
* Bus vehicle kilometres per urban hectare	v.km/ha
* Minibus vehicle kilometres per urban hectare	v.km/ha
* Tram wagon kilometres per urban hectare	v.km/ha
* Light rail wagon kilometres per urban hectare	v.km/ha
* Metro wagon kilometres per urban hectare	v.km/ha
* Suburban rail wagon kilometres per urban hectare	v.km/ha
* Heavy rail wagon kilometres per urban hectare	v.km/ha
Total public transport seat kilometres of service per capita	seat km/person
* Bus seat kilometres per capita	seat km/person
* Minibus seat kilometres per capita	seat km/person
* Tram seat kilometres per capita	seat km/person
* Light rail seat kilometres per capita	seat km/person
* Metro seat kilometres per capita	seat km/person
* Suburban rail seat kilometres per capita	seat km/person
* Heavy rail seat kilometres per capita	seat km/person
Overall average speed of public transport	km/h
* Average speed of buses	km/h
* Average speed of minibuses	km/h
* Average speed of trams	km/h

## Annexe V - Composition de la base

---

* Average speed of light rail	km/h
* Average speed of metro	km/h
* Average speed of suburban rail	km/h
* Average speed of heavy rail	km/h
<b>Mobility Indicators</b>	
<b>Overall mobility</b>	
Daily trips by foot per capita	trips/person
Daily trips by mechanized, non motorised modes per capita	trips/person
Daily public transport trips per capita	trips/person
Daily private transport trips per capita	trips/person
Total daily trips per capita	trips/person
<b>Mode split of all trips</b>	
* Percentage of non motorised modes over all trips	%
* Percentage of motorised public modes over all trips	%
* Percentage of motorised private modes over all trips	%
<b>Mode split of mechanised trips</b>	
* Percentage of mechanised, non motorised modes over mechanised trips	%
* Percentage of motorised public modes over mechanised trips	%
* motorised private modes	%
Overall average trip distance	km
Overall average trip distance by car	km
Overall average trip distance by public transport	km
Average distance of mechanised trips	km
Overall average distance of the journey-to-work	km
Average distance of the journey-to-work by mechanised modes	km
Average time of a car trip	minutes
Average time of a public transport trip	minutes
<b>Private Mobility Indicators [cars and motorcycles]</b>	
Passenger car kilometres per capita	v.km/person
Motor cycle kilometres per capita	v.km/person
Total private passenger vehicle kilometres per capita	v.km/person
Passenger car passenger kilometres per capita	p.km/person
Motor cycle passenger kilometres per capita	p.km/person
Total private passenger kilometres per capita	p.km/person
<b>Private Mobility Indicators [taxis and shared taxis]</b>	
Taxi passenger kilometres per capita	p.km/person
Shared taxi passenger kilometres per capita	p.km/person
Taxi trips per capita	trips/person
Shared taxi trips per capita	trips/person
<b>Public Transport Mobility Indicators</b>	
Total public transport boardings per capita	boardings/person
* Bus boardings per capita	boardings/person
* Minibus boardings per capita	boardings/person
* Tram boardings per capita	boardings/person
* Light rail boardings per capita	boardings/person
* Metro boardings per capita	boardings/person
* Suburban rail boardings per capita	boardings/person
* Heavy rail boardings per capita	boardings/person
Total public transport passenger kilometres per capita	p.km/person
* Bus passenger kilometres per capita	p.km/person
* Minibus passenger kilometres per capita	p.km/person
* Tram passenger kilometres per capita	p.km/person

## Annexe V - Composition de la base

---

* Light rail passenger kilometres per capita	p.km/person
* Metro passenger kilometres per capita	p.km/person
* Suburban rail passenger kilometres per capita	p.km/person
* Heavy rail passenger kilometres per capita	p.km/person
User cost of transport	
Average user cost of a car trip	10-4 % per capita GDP/trip
Average user cost of a public transport trip	10-4 % per capita GDP/trip
Price of fuel per km	10-5 % per capita GDP/km
User cost of private transport per passenger kilometre	10-5 % per capita GDP/km
User cost of public transport per passenger kilometre	10-5 % per capita GDP/km
Maximum charge for on-street parking in the CBD	10-4 % per capita GDP/h
Maximum charge for off-street parking in the CBD	10-4 % per capita GDP/h
Average of the maximum parking charges in the CBD	10-4 % per capita GDP/h
Fine for parking in no parking zone	% per capita GDP
Fine for obstructing public transport	% per capita GDP
Fine for exceeding parking time in a paying parking place	% per capita GDP
Public transport productivity	
Overall public transport vehicle occupancy	persons/unit
* Bus vehicle occupancy	persons/unit
* Minibus vehicle occupancy	persons/unit
* Tram wagon occupancy	persons/unit
* Light rail wagon occupancy	persons/unit
* Metro wagon occupancy	persons/unit
* Suburban rail wagon occupancy	persons/unit
* Heavy rail wagon occupancy	persons/unit
Overall public transport seat occupancy	persons/seat
* Bus seat occupancy	persons/seat
* Minibus seat occupancy	persons/seat
* Tram seat occupancy	persons/seat
* Light rail seat occupancy	persons/seat
* Metro seat occupancy	persons/seat
* Suburban rail seat occupancy	persons/seat
* Heavy rail seat occupancy	persons/seat
Public transport operating cost recovery	%
Average public transport farebox revenue per boarding	USD/boarding
Average public transport farebox revenue per passenger kilometre	USD/pass. km
Average public transport farebox revenue per vehicle kilometre	USD/v.km
Transport Financial Cost	
Public Transport Cost	
Percentage of metropolitan GDP spent on public transport investment	%
Public transport investment per capita	USD/person
Public transport operating cost per vehicle kilometre	USD/v.km
Public transport operating cost per passenger kilometre	USD/pass. km
Public transport operating cost per capita	USD/person
Percentage of metropolitan GDP spent on public transport operating costs	%
Private Transport Cost	
Percentage of metropolitan GDP spent on road investment	%
Road investment per capita	USD/person
Annual road investment per kilometre of road	USD/km
Private transport operating cost per vehicle kilometre	USD/v.km
Private transport operating cost per passenger kilometre	USD/pass. km
Private transport operating cost per capita	USD/person

## Annexe V - Composition de la base

Percentage of metropolitan GDP spent on private transport operating costs	%
Overall Transport Cost	
Overall transport cost per passenger kilometre	USD/pass. km
Overall private transport cost per passenger kilometre	USD/pass. km
Overall public transport cost per passenger kilometre	USD/pass. km
Total passenger transport cost per capita	USD/person
Total private passenger transport cost per capita	USD/person
Total public passenger transport cost per capita	USD/person
Total passenger transport cost as percentage of metropolitan GDP	%
Total private passenger transport cost as percentage of metropolitan GDP	%
Total public passenger transport cost as percentage of metropolitan GDP	%
Transport Externalities Indicators	
Transport Energy Indicators	
Private passenger transport energy use per capita	MJ/person
Public transport energy use per capita	MJ/person
Total transport energy use per capita	MJ/person
Energy use per private passenger vehicle kilometre	MJ/km
Energy use per public transport vehicle kilometre	MJ/km
* Energy use per bus vehicle kilometre	MJ/km
* Energy use per minibus vehicle kilometre	MJ/km
* Energy use per tram wagon kilometre	MJ/km
* Energy use per light rail wagon kilometre	MJ/km
* Energy use per metro wagon kilometre	MJ/km
* Energy use per suburban rail wagon kilometre	MJ/km
* Energy use per heavy rail wagon kilometre	MJ/km
Energy use per private passenger kilometre	MJ/p.km
Energy use per public transport passenger kilometre	MJ/p.km
* Energy use per bus passenger kilometre	MJ/p.km
* Energy use per minibus passenger kilometre	MJ/p.km
* Energy use per tram passenger kilometre	MJ/p.km
* Energy use per light rail passenger kilometre	MJ/p.km
* Energy use per metro passenger kilometre	MJ/p.km
* Energy use per suburban rail passenger kilometre	MJ/p.km
* Energy use per heavy rail passenger kilometre	MJ/p.km
Overall energy use per passenger kilometre	MJ/p.km
Air Pollution Indicators	
Total emissions per capita	
* Emissions of CO per capita	kg/person
* Emissions of SO <sub>2</sub> per capita	kg/person
* Emissions of VHC per capita	kg/person
* Emissions of NO <sub>x</sub> per capita	kg/person
Total emissions per urban hectare	kg/ha
Total emissions per total hectare	kg/ha
Transport Fatalities Indicators	
Total transport deaths per million people	deaths/10 <sup>6</sup> persons
Total transport deaths per billion vehicle kilometres	deaths/10 <sup>9</sup> v.km
Total transport deaths per billion passenger kilometres	deaths/10 <sup>9</sup> p.km
Public/Private Transport Balance Indicators	
Proportion of total motorised passenger kilometres on public transport	%
Ratio of public versus private transport speeds	
Ratio of annual investment in public transport versus private transport infrastructure	

## Annexe V - Composition de la base

---

Ratio of segregated public transport infrastructure versus expressways  
Ratio of public versus private transport energy use per passenger kilometre  
Ratio of public vs private transport total cost  
Ratio of public versus private transport user cost per passenger kilometre

## Annexe VI - Détermination des éléments constituant la valeur du temps (Chapitre 5, p. 244)

Le programme d'optimisation du modèle d'allocation du temps aux activités de Jara-Diaz (2003) est le suivant :

$$\underset{X,T}{\text{Max}} U(X,T)$$

sous les contraintes :

$$wT_w - \sum_{i=1}^G P_i X_i \geq 0 \quad (\lambda)$$

$$\tau - \sum_{j=1}^A T_j = 0 \quad (\mu)$$

$$T_j - f_j(X) \geq 0 \quad (\kappa_j) \quad \forall j = 1, \dots, A$$

$$X_i - g_i(T) \geq 0 \quad (\psi_i) \quad \forall i = 1, \dots, G$$

Afin de tenir compte des relations entre les temps d'activités, nous proposons l'introduction de la contrainte technique suivante :

$$T_j - h_j(T_j) \geq 0 \quad (\varphi_j)$$

où  $T_j$  est le vecteur de durées d'activités différentes de  $j$ . L'hypothèse simplificatrice est faite que la durée minimale requise d'une activité est uniquement dépendante des durées des autres activités. La durée  $T_j$  n'a pas d'effet sur son niveau minimum.

Sans cette hypothèse, le Lagrangien a pour expression :

$$\begin{aligned} \text{Lagrangien} = & U(X,T) - \lambda \left( wT_w - \sum_{i=1}^G p_i X_i \right) - \mu \left( \tau - \sum_{j=1}^A T_j \right) \\ & + \sum_{j=1}^A \kappa_j (T_j - f_j(X)) + \sum_{i=1}^G \psi_i (X_i - g_i(T)) + \sum_{j \neq k}^A \varphi_j (T_j - h_j(T)) \\ & + \varphi_k (1 - h_k(T)) \end{aligned}$$

où le dernier terme correspond à l'effet de la contrainte liant la durée d'une activité avec l'effet de sa réduction de son temps minimum requis. L'hypothèse émise élimine ce dernier terme.

Ainsi, les éléments suivants de la valeur du temps découlent aisément des conditions de premier ordre de ce programme d'optimisation :

$$\frac{\mu}{\lambda} = w + \frac{\partial U / \partial T_w}{\lambda} - \frac{1}{\lambda} \sum_{i=1}^G \psi_i \frac{\partial g_i}{\partial T_w} - \frac{1}{\lambda} \sum_{j=1}^A \varphi_j \frac{\partial h_j}{\partial T_w}$$

$$\frac{\kappa_k}{\lambda} = \frac{\mu}{\lambda} - \frac{\partial U / \partial T_k}{\lambda} + \frac{1}{\lambda} \sum_{i=1}^G \psi_i \frac{\partial g_i}{\partial T_k} + \frac{1}{\lambda} \sum_{j=1}^A \varphi_j \frac{\partial h_j}{\partial T_k}$$

$$\frac{\kappa_k}{\lambda} = w + \frac{\partial U / \partial T_w}{\lambda} - \frac{\partial U / \partial T_k}{\lambda} + \sum_{i=1}^G \frac{\psi_i}{\lambda} \left( \frac{\partial g_i}{\partial T_k} - \frac{\partial g_i}{\partial T_w} \right) + \sum_{j=1}^A \frac{\varphi_j}{\lambda} \left( \frac{\partial h_j}{\partial T_k} - \frac{\partial h_j}{\partial T_w} \right)$$

$$\frac{\psi_i}{\lambda} = P_i - \frac{\partial U / \partial X_i}{\lambda} + \frac{1}{\lambda} \sum_{j=1}^A \kappa_j \left( \frac{\partial f_j}{\partial X_i} \right)$$

$$\frac{\varphi_k}{\lambda} = \frac{\mu}{\lambda} + \frac{\partial U / \partial T_k}{\lambda} + \frac{1}{\lambda} \sum_{i=1}^G \psi_i \frac{\partial g_i}{\partial T_k} + \frac{1}{\lambda} \sum_{j=1}^A \varphi_j \frac{\partial h_j}{\partial T_k}$$

Annexe VII - Récapitulatif des modèles d'allocation  
des temps

### Synthèse des modèles d'allocation des temps

	<b>Golob et al. (1981) <i>UMOT</i></b>	<b>Kraan (1997)</b>	<b>Bhat et Misra (1999)</b>
<b>Utilité</b>	$U(x, c, L) = \phi(x) + \varphi(c) + \xi(L)$	$U(T_i, d_i, f_i, \forall i; T_H, G) = \sum_i [Z_i(T_i, d_i, f_i) - R_i(d_i)] + Z_H(T_H) + Z_G(G)$	$U(s_1, s_2, s_3, s_4) = \sum_{i=1}^4 e^{\beta_i X} \ln(s_i)$
<b>Equation intermédiaire</b>	$U(x, c, L) = \sum_{j=1}^m a_j \log(x_j) + b_1 \log\left(M - \sum_{j=1}^m c_j x_j\right) + b_2 \log\left(\bar{T} - \sum_{j=1}^m \frac{x_j}{v_j}\right)$	$U(T_i, d_i, f_i, \forall i; T_H, G) = \sum_i \left[ T_i^{\beta_i} d_i^{\gamma_i} f_i^{\rho_i} - \zeta_t \frac{d_i}{v_i} - \zeta_c c_i^d d_i \right] + T_H^g + G^z$	$s_i^* = \frac{e^{\beta_i X}}{\sum_{j=1}^4 e^{\beta_j X}}$
<b>Contrainte budgétaire (<math>\lambda</math>)</b>	$p_x x + p_c c \leq M$	$\sum_i \left[ T_i + \frac{d_i}{v_i} \right] + T_H = T_{tot}$	
<b>Contrainte temporelle (<math>\mu</math>)</b>	$t_x x + L \leq \bar{T}$	$\sum_i c_i(T_i, d_i, f_i) + G = Y$ $T_i, d_i, f_i \geq 0 \quad \forall i$ $T_H, G \geq 0$	$\sum_{i=1}^4 s_i = 1$ $s_i > 0$
<b>Système estimé</b>	$x_j^* = \frac{a_j}{\frac{b_1 c_j}{M} + \frac{b_2}{v_j \bar{T}}}$	$\left\{ \begin{array}{l} \ln(T_H) = \frac{\ln\left[\frac{\beta}{g}\right] + \rho \ln\left[\frac{\rho v}{\beta d}\right]}{g-1} + \frac{\beta + \rho - 1}{g-1} \ln(T) \\ f = \frac{\rho v}{\beta d} T \\ t_t = \frac{\rho}{\beta} T \end{array} \right.$	$\ln\left(\frac{s_i^*}{s_1^*}\right) = \beta_i' X, \quad i=1, 2, 3, 4$

### Synthèse des modèles d'allocation des temps (suite)

	<b>Kitamura (1984)</b>	<b>Kockelman (2001)</b>
<b>Utilité</b>	$U(t_1, \dots, t_i, x_i) = \sum_{i=1}^I \xi_i V_i(t_i, x_i)$	$\max_{A, T, Z} u(A, T, tA, Z)$
<b>Equation intermédiaire</b>	$V_i(t_i, x_i) = \gamma_i f_i(x_i) \ln t_i$ <p style="text-align: center;">et</p> $f_i(x_i) = \prod_{k=1}^{K_i} x_{ik}^{\beta_{ik}}$	<p style="text-align: center;">Fonction d'utilité indirecte :</p> $v = \alpha_0 + \sum_i \alpha_i \ln(t_i) + \frac{1}{2} \sum_{i,j} \beta_{ij} \ln(t_i) \ln(t_j) + \sum_i \gamma_i \ln(T_d) \ln(t_i)$ $+ \sum_i \gamma_{iY} \ln(Y) \ln(t_i) + \gamma_i \ln(Y) \ln(T_d)$
<b>Contrainte budgétaire (<math>\lambda</math>)</b>	-	$P_A A + P_t A + P_Z Z \leq Y + w T_w$
<b>Contrainte temporelle (<math>\mu</math>)</b>	$\sum_{i=1}^I t_i = T$ $t_i \geq 0, \quad i = 1, \dots, I$	$\sum_i T_i + tA = T, \quad i = 1, \dots, n$ $A, tA, Z \geq 0$
<b>Système estimé</b>	$\ln \left( \frac{t_i^*}{t_j^*} \right) = \sum_{k=1}^{k_i} \beta_{1k} \ln x_{1k} + \eta_1$ $\ln \left( \frac{t_{j-1}^*}{t_j^*} \right) = \sum_{k=1}^{k_{j-1}} \beta_{j-1,k} \ln x_{j-1,k} + \eta_{j-1}$ $\ln \left( \frac{t_{j+1}^*}{t_j^*} \right) = \sum_{k=1}^{k_{j+1}} \beta_{j+1,k} \ln x_{j+1,k} + \eta_{j+1}$ <p style="text-align: center;">...</p> $\ln \left( \frac{t_i^*}{t_j^*} \right) = \sum_{k=1}^{k_i} \beta_{ik} \ln x_{ik} + \eta_i$	$X_i^* = \frac{-\frac{1}{t_i} \left( \alpha_i + \sum_j \beta_{ij} \ln(t_j) + \gamma_{iY} \ln(Y) + \gamma_{iT} \ln(T_d) \right)}{\frac{1}{T_d} \left( \sum_j \gamma_{jT} \ln(t_j) + \gamma_{TY} \ln(Y) \right)}$

### Synthèse des modèles d'allocation des temps (suite)

	<b>Chen et Mokhtarian (2005)</b>
<b>Utilité</b>	$\max U(t_w, t_m, t_d, t_t, G)$
<b>Equation intermédiaire</b>	<p style="text-align: center;">Fonction de demande :</p> $\ln c(u, p) = \alpha_0 + \sum_k \alpha_k \ln p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj}^* \ln p_k \ln p_j + \mu \beta_0 \prod_k p_k^{\beta_k}$ $\sum_i \alpha_i = 1, \quad \sum_j \gamma_{kj}^* = \sum_k \gamma_{kj}^* = \sum_j \beta_j = 0$
<b>Contrainte budgétaire (<math>\lambda</math>)</b>	$c_m t_m + c_d t_d + c_t t_t + G = w t_w + Y$
<b>Contrainte temporelle (<math>\mu</math>)</b>	$t_w + t_m + t_d + t_t = \tau$ $t_t \geq b_w t_w + b_m t_m + b_d t_d$ $b_w, b_m, b_d \geq 0$
<b>Système estimé</b>	$\frac{p_w t_w}{Y} = \alpha_w + \gamma_{ww} \ln p_w + \gamma_{wm} \ln p_m + \gamma_{wd} \ln p_d + \beta_w \ln(Y/P)$ $\frac{p_w t_m}{Y} = \alpha_m + \gamma_{mw} \ln p_w + \gamma_{mm} \ln p_m + \gamma_{md} \ln p_d + \beta_m \ln(Y/P)$ $\frac{p_w t_d}{Y} = \alpha_d + \gamma_{dw} \ln p_w + \gamma_{dm} \ln p_m + \gamma_{dd} \ln p_d + \beta_d \ln(Y/P)$ $\frac{G}{Y} = \alpha_G + \gamma_{Gw} \ln p_w + \gamma_{Gm} \ln p_m + \gamma_{Gd} \ln p_d + \beta_G \ln(Y/P)$ $\ln P = \alpha_0 + \alpha_w \ln p_w + \alpha_m \ln p_m + \alpha_d \ln p_d + \frac{1}{2} \gamma_{ww} (\ln p_w)^2 + \frac{1}{2} \gamma_{mw} \ln p_m \ln p_w$ $+ \frac{1}{2} \gamma_{dw} \ln p_d \ln p_w + \frac{1}{2} \gamma_{wm} \ln p_w \ln p_m + \frac{1}{2} \gamma_{mm} (\ln p_m)^2 + \frac{1}{2} \gamma_{dm} \ln p_d \ln p_m$ $+ \frac{1}{2} \gamma_{wd} \ln p_w \ln p_d + \frac{1}{2} \gamma_{md} \ln p_m \ln p_d + \frac{1}{2} \gamma_{dd} (\ln p_d)^2$

$J$  : les modes de transport  
 $I$  : les activités  
 $K$  : les variables exogènes  
 $x$  : quantité de transport (distance parcourue)  
 $c$  : consommation non-transport  
 $L$  : temps de loisir  
 $M$  : budget monétaire  
 $Y$  : revenu non-salarial  
 $T$  : temps disponible hors travail net du temps de la consommation  
 $p_x$  : prix du bien transport  
 $p_c$  : prix des biens de consommation  $c$   
 $t_x$  : temps de transport par unité de transport  $x$   
 $a_i$  : coefficient d'attractivité du mode  $i$   
 $b_1$  et  $b_2$  : pondérations des utilités de la consommation et du loisir dans la fonction d'utilité totale

$T_i$  : durée d'activité  $i$   
 $d_i$  : distance parcourue pour l'activité  $i$   
 $f_i$  : nombre de réalisation de l'activité  $i$   
 $T_H$  : temps passé à domicile  
 $T_{tot}$  : temps disponible hors-travail  
 $G$  : dépense en biens non-transport  
 $Z_i$  : utilité de l'activité  $i$   
 $R_i$  : fonction de résistance du transport  
 $Z_H$  : utilité des activités à domicile  
 $Z_G$  : utilité de la consommation  
 $\beta_i, \gamma_i, \rho_i, \vartheta, \chi$  : paramètres à estimer  
 $\zeta_t/\zeta_c$  : valeur du temps  
 $v_i$  : vitesse moyenne  
 $c_i d$  : coût monétaire de la distance  $d_i$

$s_i$  : part du temps discrétionnaire alloué à l'activité  $i$   
 $X$  : matrice des variables exogènes

$t_j$  : durée de l'activité  $j$   
 $x_j$  : vecteur des variables exogènes  
 $V_j$  : utilité de l'activité  $j$   
 $\xi_j$  : variables aléatoires positives i.i.d.  
 $T$  : temps total disponible  
 $f_j$  : fonction des variables  $x_j$   
 $K_j$  : nombre de variables exogènes relatives à l'activité  $j$   
 $\beta_{ik}$  : paramètres à estimer  
 $\eta_j$  : variable aléatoire telle que :  $\xi_j = \exp(\eta_j)$

$A$  : le vecteur des activités hors-domicile  
 $T$  : le vecteur des temps consacrés aux activités  
 $tA$  : le vecteur des temps de transport associés à chaque activité  
 $Z$  : le vecteur de biens de consommation  
 $G$  : dépense en biens non-transport

$t_w$  : temps de travail  
 $t_m$  : temps obligatoire  
 $t_d$  : temps discrétionnaire  
 $t_i$  : temps de transport  
 $c_m$  : coût horaire des activités obligatoires  
 $c_d$  : coût horaire des activités discrétionnaires  
 $c_i$  : coût horaire du transport  
 $G$  : dépense en biens non-transport  
 $b_w$  : temps de transport par unité de temps de travail  
 $b_m$  : temps de transport par unité de temps obligatoire  
 $b_d$  : temps de transport par unité de temps discrétionnaire  
 $p_w$  : *full price* du travail  
 $p_m$  : *full price* des activités obligatoires  
 $p_d$  : *full price* des activités discrétionnaires  
 $\alpha_0, \alpha_k, \gamma_{kj}^*, \mu, \beta_0, \beta_k$  sont les paramètres

Annexe VIII - Récapitulatif des modèles de durées  
appliqués à l'analyse des programmes d'activités

### Synthèse des modèles de durées appliqués à l'analyse des programmes d'activités

Etude	Durée modélisée	Méthode	Résultats
<i>Durée à domicile entre deux activités</i>			
Hamed et Mannering (1993)	Durée à domicile après le travail	Weibull	La probabilité de poursuivre une activité hors domicile est affectée par les caractéristiques socio-économiques et l'horaire d'arrivée au domicile
Mannering F., E. Murakami et S.G. Kim (1994)	Durée à domicile entre deux activités hors-domicile	Modèles de Cox test de stabilité des coefficients sur deux années	Instabilité du modèle ; Effet des variables relatives aux habitudes et aux horaires de retour à domicile
<i>Durée des activités</i>			
Niemeier et Morita (1996)	Durées d'activités d'achats, de loisirs et d'affaires personnelles en distinguant hommes et femmes	Modèle non-paramétrique et semi-paramétrique de Cox	Différence d'allocation significative pour l'achat ; Les femmes ont des durées d'achat plus longues ; L'inscription des achats sur un trajet pour le travail allonge la durée d'achat.
Bhat (1996a)	Durée d'achat sur le trajet travail-domicile	Modèle non-paramétrique à hasards proportionnels tenant compte d'une hétérogénéité inobservée non-paramétrique	Les biais des estimations paramétriques de la dépendance temporelle et de l'hétérogénéité inobservée ; Effet du statut professionnel et du mode de transport.
Kitamura et al. (1997a)	durées des activités pour les actifs et les inactifs (illustration des résultats pour les loisirs récréatif, les loisirs sociaux, les durées d'activités sur le trajet travail-domicile)	Modèle paramétrique Weibull	Dépendance des durées des épisodes d'activité par rapport aux activités passées ; Influence des horaires de réalisation et des attributs socio-économiques

### Synthèse des modèles de durées appliqués à l'analyse des programmes d'activités (suite)

Etude	Durée modélisée	Méthode	Résultats
<i>Durée des activités (suite)</i>			
Yee et Niemeier (2000)	Durées de visites, d'activité sociales, de loisir et d'affaire personnelles observées en quatre dates de 1989-1993	Modèle semi-paramétrique de Cox	Instabilité des coefficients ; Fort effet propre à l'année considérée.
Timmermans et al. (2002)	Durées de travail, d'études, d'achat, de loisirs récréatifs et de loisirs sociaux	Modèle semi-paramétrique de Cox	Tendances communes aux différentes régions ; Les types de ménages et les jours de réalisation des activités sont plus influentes que les contextes spatiaux ou les systèmes de transport ; Différences entre les régions plus faibles que les différences intra-zones
Schwanen (2004)	Durée d'achat selon sa position dans le programme d'activités	Modèle semi-paramétrique de Cox	Effet significatif de l'offre de commerces plus large et diversifiée des zones urbanisées
Srivinasan et Bhat (2005)	Stratégie du ménage pour la participation et la durée des activités de maintenance hors-domicile	Modèle de durées joint à un modèle logit mixte estimés simultanément	Spécialisation des tâches, effets des temps obligatoires, effets des variables socio-économiques.
<i>Durées avant réalisation – choix d'horaire</i>			
Wang (1996)	Durée avant l'engagement d'une des 15 activités d'un emploi du temps	Modèle paramétrique Weibull	Estimation des horaires préférés pour l'engagement des activités
Bhat et Steed (2002)	Durée avant l'engagement d'une activité d'achat quotidiens	Modèle non-paramétrique tenant compte de l'hétérogénéité inobservée gamma et autorisant les effets variables dans le temps	Effets des variables socio-économiques ; Absence d'effet des coûts et des temps de trajet ; Présence d'hétérogénéité

### Synthèse des modèles de durées appliqués à l'analyse des programmes d'activités (suite)

Etude	Durée modélisée	Méthode	Résultats
<i>Durées d'activités et les probabilités de transition entre activités</i>			
Ettema, Borgers et Timmermans (1995)	Durée de l'activité en cours, choix de la prochaine activité et leur dépendance mutuelle	Modèle de durées à risques concurrents et durée de vie accélérée	Différences significatives en termes de critères d'organisation des différents types d'activités
Ma et Goulias (1998)	Choix d'horaire ; durées d'activités ; temps de transport par type d'activité	Modèle à risques concurrents ; modèles à durées de vie accélérée ; modèles à durée de vie accélérée itérés	Les temps de transport sont endogènes aux durées d'activités
Bhat (1996b)	Durées d'activités durant le trajet travail-domicile	Modèle généralisé non-paramétrique à hasards proportionnels multiples	Endogénéité entre le type et la durée de l'activité ; effet des variables socio-économiques et du travail
Popkowski Leszczys et Timmermans (2002)	Probabilité de transition et durée des activités	Modèle à risques (non)concurrents (in)conditionnels	Le modèle à risques concurrents conditionnels est le plus performant ; la durée et la nature de l'activité précédente affecte le choix du type d'activité et l'horaire de réalisation ; les variables socio-économiques affectent les probabilités de transition entre activités.
<i>Durée inter-épisode</i>			
Schönfelder et Axhausen (2001)	Durée inter-achats	Modèle paramétrique Weibull et modèle semi-paramétrique de Cox	Effet des variables socio-économiques ; différences entre les deux villes
Bhat, Sivakumar et Axhausen (2003)	Durées inter-épisodes d'achat non-relatifs à la maintenance du ménage	Modèle non-paramétrique à hasards proportionnels tenant compte de l'hétérogénéité inobservée	Impact des technologies de communication sur la durée interachats ; hétérogénéité de l'usage des technologies de communication

### Synthèse des modèles de durées appliqués à l'analyse des programmes d'activités (suite)

Etude	Durée modélisée	Méthode	Résultats
<i>Durée inter-épisode (suite)</i>			
Bhat, Frusti, Schönfelder, Axhausen (2004a)	Durée inter-achats	Modèle non-paramétrique à hasards proportionnels tenant compte de l'hétérogénéité inobservée	Tendance hebdomadaire ; hasard non-monotone ; les statuts professionnels de l'individu et de son conjoint, le mode de transport utilisé, et le comportement d'enchaînement des déplacements sont influents
Bhat, Srinivasan et Axhausen (2005)	Durées inter-épisodes pour différents types d'activités (deux activités d'achats de maintenance ou non et trois activités non-achat : loisir, affaires personnelles et activité sociale)	Modèle généralisé non-paramétrique à hasards proportionnels multiples tenant compte de l'hétérogénéité inobservée	Probabilité de participation aux activités d'achat est croissante avec le temps ; participation aux activités non-achat sont fortement marquées par un rythme hebdomadaire ; les attributs de l'individu, du conjoint et du ménage, la localisation résidentielle, le jour de semaine ont une forte influence sur la durée inter-épisode ; le taux de motorisation et le lien entre la localisation et le système de transport n'ont pas d'effet sur les taux de participations aux activités
<i>Durée entre la planification et l'exécution</i>			
Mohammadian et Doherty (2004)	Durée entre la planification d'une activité et son exécution	Modèles de durées paramétriques avec hétérogénéité gamma et semi-paramétriques	Effets des attributs et des mesures de flexibilité des activités, des caractéristiques individuelles et des caractéristiques du programme d'activités total hebdomadaire

## Annexe IX - Présentation des modèles de durées

### I. Principes généraux

L'application des modèles de durées en économie des transports est relativement récente. Il apparaît donc nécessaire de rappeler les éléments principaux de cette méthodologie pour une meilleure compréhension des hypothèses du modèle et des résultats de son application aux BTT. Le lecteur souhaitant développer les notions abordées et étudier d'autres méthodes d'estimation des modèles de durées pourra se référer aux ouvrages suivants : Kalbfleisch et Prentice (1980), Lawless (2003), Hosmer et Lemeshow (1999), Allison (1995).

Avant de présenter dans le détail les modèles de durées, examinons tout d'abord les hypothèses à la base du modèle. L'objectif est la modélisation et l'étude de la durée quotidienne consacrée aux transports ainsi que de la probabilité d'interruption de ce processus au cours du temps : sa dynamique temporelle.

Définissons  $T$ , la variable aléatoire réelle positive, représentant le temps qu'un individu consacre à ses transports quotidiennement. Nous supposons que  $T$  est continue<sup>92</sup> et que les individus sont homogènes par rapport à leurs comportements d'allocation de temps aux transports. La distribution de  $T$  est caractérisée par la fonction de densité,  $f(t)$  et la fonction de répartition,  $F(t)$  définies par :

$$f(t) = \lim_{\Delta \rightarrow 0^+} \frac{P(t \leq T < t + \Delta)}{\Delta} \quad (eq. 1)$$

et

$$F(t) = P(T \leq t) = \int_0^t f(u) du \quad (eq. 2)$$

---

<sup>92</sup> La continuité de la variable modélisée est une hypothèse critiquable pour l'analyse du temps de transport quotidien. En effet, le budget-temps de transport est considéré comme la somme des durées des déplacements réalisés dans une journée. Il est dépendant du nombre de déplacements et n'est donc pas une variable continue dans le processus décisionnel de l'individu. Cependant, le grand nombre d'observations dont nous disposons ainsi que le grand nombre de valeurs prises par les temps de transport, nous permettent de soutenir cette hypothèse récurrente dans l'ensemble des modèles de temps de transport.

Ainsi la probabilité d'interruption du processus dans un intervalle de temps  $\Delta$ , est donnée par :  $f(t) \cdot \Delta$ . Et la probabilité que l'interruption survienne avant la date  $t$  est :  $F(t)$ .

La probabilité complémentaire de  $F(t)$  est donnée par la fonction de survie ou la fonction d'endurance (Bhat, 2000), notée  $S(t)$ . Elle correspond à la probabilité que le processus dure jusqu'à la date  $t$  :

$$S(t) = \Pr[T > t] = 1 - F(t) = \int_t^{\infty} f(u) du \quad (eq. 3)$$

Les modèles de durées estiment les probabilités d'interruption sur un intervalle de temps infinitésimal  $\Delta$  après la date  $t$ , sachant que le processus a duré jusqu'à une date donnée  $t$ . Cette probabilité conditionnelle est notée  $h(t)$  et est qualifiée de fonction de hasard ou encore de taux de hasard. Elle est définie par :

$$h(t) = \lim_{\Delta \rightarrow 0^+} \frac{P(t \leq T < t + \Delta | T > t)}{\Delta} \quad (eq. 4)^{93}$$

Enfin, le hasard intégré est défini par :

$$H(t) = \int_0^t h(u) du = -(\ln S(t)) \quad (eq. 5)$$

Il ne possède pas d'interprétation directe.

Ces cinq fonctions caractérisent la distribution de la durée,  $T$ , du processus et la variation de sa probabilité d'interruption au fil du temps.

Le hasard  $h(t)$  peut être transformé par application de la propriété des probabilités conditionnelles suivante :

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \quad (eq. 6)$$

où dans notre cas  $P(B|A) = 1$ , car  $P(t \leq T \leq t + \Delta | T > t) = 1$

Ainsi :

---

<sup>93</sup>  $P(A/B)$  note la probabilité conditionnelle de l'évènement A sachant l'évènement B.

$$\begin{aligned}
 h(t) &= \lim_{\Delta \rightarrow 0^+} \frac{P(t \leq T < t + \Delta | T > t)}{\Delta} \\
 &= \lim_{\Delta \rightarrow 0^+} \frac{P(t \leq T < t + \Delta)}{\Delta} \cdot \frac{1}{P(T > t)} \\
 &= \lim_{\Delta \rightarrow 0^+} \frac{F(t + \Delta) - F(t)}{(1 - F(t)) \cdot \Delta}
 \end{aligned}$$

Le hasard se définit comme le rapport des fonctions de densité et de survie :

$$h(t) = \frac{f(t)}{S(t)} \quad (\text{eq. 7})$$

De plus, le lien entre le hasard et la survie est obtenu par :

$$\begin{aligned}
 h(t) &= \frac{f(t)}{S(t)} = \frac{dF(t)/dt}{S(t)} = \frac{d[1 - S(t)]}{S(t)} = \frac{-dS(t)/dt}{S(t)} \\
 h(t) &= \frac{-d \ln S(t)}{dt} \quad (\text{eq. 8})
 \end{aligned}$$

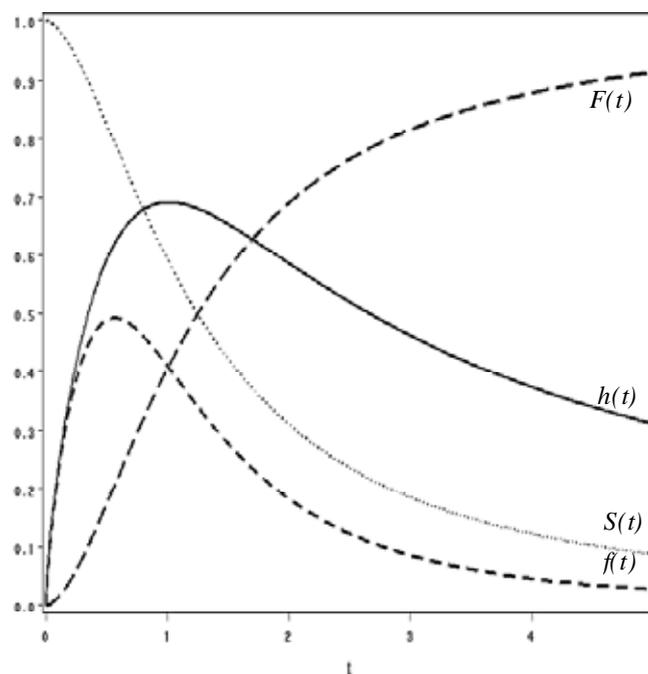
En définitive, la probabilité conditionnelle peut se déterminer par le quotient de la probabilité instantanée d'interruption à la date  $t$ , et de la probabilité de survie jusqu'à la date  $t$ . La fonction de hasard exprime quant à elle, l'opposé de la pente de la courbe de survie en valeur relative.

En conséquence, la pente de la fonction de hasard indique les variations de la pente de la survie. Les variations du hasard correspondent aux modifications du rythme de la survie, l'accélération ou la décélération du processus. La forme de la fonction de hasard a donc d'importantes implications sur la représentation de la dynamique temporelle du processus (Hensher et Mannering, 1994). Par exemple, avec une pente croissante, le hasard décrit un processus dont la fonction de survie décroît de plus en plus vite. La figure 1 illustre un exemple de formes liant ces différentes fonctions lorsque le hasard est supposé non-monotone.

Les courbes  $F$  et  $f$  illustrent les fonctions de répartition et de densité. Toutes deux sont des probabilités non-conditionnelles. Les modèles de durées se concentrent plus particulièrement sur les fonctions de survie et de hasard. La probabilité d'atteindre une date donnée  $t$  est illustrée par la courbe de survie,  $S$ . Enfin, la pente de la courbe de hasard,  $h$ , nous informe directement sur les variations du rythme de décroissance de la survie. La probabilité conditionnelle d'interruption est dépendante du temps écoulé. Le hasard représenté dans la

figure 1 admet un renversement de sa pente. Ainsi, cette probabilité conditionnelle est croissante avec le temps écoulé, puis décroissante. L'interruption est donc de plus en plus probable au départ du processus, puis de moins en moins probable pour des durées importantes. L'information sur cette dynamique temporelle du processus est directement révélée par l'étude de la pente de la courbe de hasard.

**figure 1 : Exemples de formes des fonctions de densité, de distribution, de hasard et de survie**



La pente du hasard et la forme de la fonction de survie peuvent être étudiées au travers de trois approches : paramétrique, non-paramétrique et semi-paramétrique. Dans les méthodes semi-paramétriques et paramétriques, les fonctions de distribution des durées sont supposées dérivées de distributions connues. Les méthodes paramétriques reposent sur des hypothèses fortes, et grâce à cela, permettent des interprétations et surtout des prédictions robustes. Un des objectifs de l'estimation sera alors de déterminer quelle est la distribution la plus adaptée à la représentation de l'échantillon et de déterminer ses paramètres. A l'opposé, dans les méthodes non-paramétriques, aucune hypothèse n'est formulée sur la distribution *a priori* des durées. Elles décrivent les processus étudiés directement à partir des observations, mais produisent moins aisément des prédictions.

## II. L'estimation non-paramétrique – inférence non-paramétrique

### 1. L'estimateur de Kaplan-Meier

L'approche non-paramétrique se rapproche d'une analyse des statistiques descriptives des durées. La fonction de survie est estimée à l'aide de l'estimateur (KM) du produit limite de Kaplan-Meier (1958). L'estimateur KM de la survie à la date  $t_j$  est calculé comme le produit des proportions suivantes :

$$S_{KM}(t_j) = \prod_{k=1}^j \frac{n(t_k) - d(t_k)}{n(t_k)}, \quad (eq. 9)$$

où  $n(t_k)$  est la population à risque à la date  $t_k$ .  $d(t_k)$  est le nombre d'interruptions à la date  $t_k$ . La fonction de survie correspondante est alors une fonction en paliers avec une marche pour chaque temps discret correspondant à une interruption. L'estimateur KM considère les probabilités de survie à chaque palier, puis les multiplie pour obtenir un estimateur de la survie.

Lorsque le processus étudié révèle un grand nombre de temps d'interruption identiques, la définition des marches est problématique. Cette discrétisation des temps d'interruption peut notamment survenir, lorsque les temps déclarés sont arrondis, réduisant ainsi la représentation des interruptions entre eux<sup>94</sup>. Dans ce cas, les temps d'interruption doivent être regroupés dans des intervalles réguliers. La définition des marches de la fonction est donnée par le choix arbitraire des intervalles des temps d'interruption. L'estimation de la pente du hasard est réalisée sous l'hypothèse d'un hasard constant entre chaque temps discret, donc d'une distribution uniforme des interruptions dans les intervalles. Cette méthode est connue sous le nom de table de survie (*lifetable method*).

La prise en compte des observations censurées est réalisée de telle sorte qu'elles contribuent à la population à risque jusqu'à ce qu'elles en disparaissent, mais ne sont pas comptabilisées comme des interruptions. Il en résulte qu'après le dernier temps observé non-censuré, l'estimation de la survie sera considérée comme non-définie. De plus, dans les tables de survie, les populations à risque sont ajustées de la moyenne des observations censurées évaluée au centre de chaque intervalle. Sous l'hypothèse d'une distribution uniforme des

---

<sup>94</sup> C'est notamment le cas des temps de déplacements déclarés, qui sont majoritairement des multiples de 5 minutes.

observations censurées sur l'intervalle, les populations à risque sont réduites de la moitié du nombre d'observations censurées sur l'intervalle :  $c/2$ . La probabilité de survie à l'intervalle est donc donnée par :  $(n-c/2-d) / (n-c/2)$ , avec  $n$  le nombre d'observations au début de l'intervalle et  $d$  le nombre d'interruptions observées dans l'intervalle et  $c$  le nombre d'observations censurées à droite dans l'intervalle.

L'application de ces méthodes non-paramétriques produit donc un estimateur de la survie pour chaque date. Dans un second temps, l'étude de cet estimateur et de ses propriétés statistiques permet de construire l'intervalle de confiance de la survie et de déterminer pour chaque date un estimateur du hasard, des quantiles de la survie, de la survie médiane résiduelle.

## 2. Survie prédite et intervalle de confiance

De la même façon que dans les modèles classiques de régression, les intervalles de confiance des valeurs prédites renseignent sur la qualité de la prédiction. Un estimateur de la variance de la survie est obtenu par l'application de la méthode delta (Hosmer et Lemeshow, 1999). Basée sur les développements limites de Taylor, cette méthode produit une estimation de la variance d'un estimateur dont la forme diffère d'une simple somme. L'obtention de la variance de l'estimateur de Kaplan-Meier de la survie est réalisée en deux temps. Tout d'abord, la variance du logarithme de l'estimateur est déterminée. Ce dernier est :

$$\begin{aligned} \ln(\hat{S}(t)) &= \sum_{i:t_i \leq t} \ln\left(\frac{n_i - d_i}{n_i}\right) \\ &= \sum_{i:t_i \leq t} \ln(\hat{p}_i) \end{aligned} \quad (eq. 10)$$

où :  $\hat{p}_i = \frac{n_i - d_i}{n_i}$ . Notons  $\hat{q}_i = 1 - \hat{p}_i$

En considérant les observations de la population à risque à la date  $t_i$  comme des réalisations indépendantes d'une variable aléatoire Bernouilli de probabilité constante, dont  $\hat{p}_i$  est un estimateur et  $(\hat{p}_i \hat{q}_i)/n_i$  un estimateur de la variance, nous pouvons déduire un estimateur de la variance de leurs transformations dans (eq. 10).

Par application de la méthode delta<sup>95</sup>, la variance du logarithme d'une variable  $X$  de moyenne  $\mu_X$  et de variance  $\sigma_X^2$  est :

$$V[\ln(X)] \cong \frac{1}{\mu_X^2} \sigma_X^2 \quad (eq. 11)$$

Et la variance de l'exponentiel de  $X$  est :

$$V(\exp(X)) \cong (\exp(\mu_X))^2 \cdot \sigma_X^2 \quad (eq. 12)$$

D'où la variance du logarithme de l'estimateur de la survie est :

$$\begin{aligned} \hat{V}[\ln(\hat{S}(t))] &= \sum_{i:t_i \leq t} \hat{V}[\ln(\hat{p}_i)] \\ &= \sum_{i:t_i \leq t} \left[ \ln\left(\frac{n_i - d_i}{n_i}\right) \right] \end{aligned} \quad (eq. 13)$$

Et la variance de l'estimateur de la survie proposé par Greenwood (1926) est :

$$\hat{V}(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)} \quad (eq. 14)$$

Il a été démontré que l'estimateur de Kaplan-Meier est asymptotiquement distribué selon une loi normale (Andersen et al., 1993, Fleming et Harrington, 1991). Ainsi, nous pouvons déduire les bornes des intervalles de confiance de la survie estimée :

$$\hat{S}(t) \pm z_{1-\alpha/2} \hat{\sigma}[\hat{S}(t)] \quad (eq. 15)$$

où  $z_{1-\alpha/2}$  est le quantile de la distribution normale standard et  $\hat{\sigma}(\hat{S}(t))$  est l'écart type estimé de  $\hat{S}(t)$ <sup>96</sup>.

### 3. Le hasard estimé

Un estimateur du hasard au point milieu de chaque intervalle peut être déterminé par le quotient de la densité et de la survie évaluées en ce point. Soit  $t_{mi}$  le milieu de  $[t_{i-1}, t_i[$  et  $b_i$  l'étendue de cet intervalle :  $b_i = t_i - t_{i-1}$ .

Par construction de l'estimateur (eq. 9), la survie estimée en chaque point  $t_i$  est :

---

<sup>95</sup> Hosmer et Lemeshow (1999), p. 354.

<sup>96</sup> Cette définition des bornes des intervalles de confiance peut conduire à des valeurs non comprises entre 0 et 1. Dans ce cas, Kalbfleisch et Prentice (1980, p.15) ont proposé d'obtenir l'intervalle de confiance pour  $-\log S(t)$ .

$$\hat{S}(t_i) = \begin{cases} 1 & \text{si } i = 0 \\ \hat{S}(t_{i-1}) \cdot \hat{p}_i & \text{si } i > 0 \end{cases} \quad (\text{eq. 16})$$

La survie estimée au point  $t_{mi}$  est :

$$\hat{S}(t_{mi}) = \frac{\hat{S}(t_i) + \hat{S}(t_{i-1})}{2} = \frac{(1 + \hat{p}_i)\hat{S}(t_{i-1})}{2} \quad (\text{eq. 17})$$

Par ailleurs, la densité peut être estimée par :

$$\hat{f}(t_{mi}) = \frac{-[\hat{S}(t_i) - \hat{S}(t_{i-1})]}{t_i - t_{i-1}} = \frac{\hat{q}_i \hat{S}(t_{i-1})}{b_i} \quad (\text{eq. 18})$$

Le hasard estimé au point milieu de chaque intervalle est :

$$\hat{h}(t_{mi}) = \frac{\hat{f}(t_{mi})}{\hat{S}(t_{mi})} = \frac{2\hat{q}_i}{b_i(1 + \hat{p}_i)} \quad (\text{eq. 19})$$

#### 4. Les quantiles de la survie estimée

Les quantiles de la survie peuvent être estimés à leur tour. Ils sont définis par :

$$\hat{t}_p = \min \{t : \hat{S}(t) \leq (p/100)\} \quad (\text{eq. 20})$$

Un estimateur de leur variance peut être obtenu par la méthode delta (Collett, 1994, cité dans Hosmer et Lemeshow, 1999) :

$$\hat{V}(\hat{t}_p) = \frac{\hat{V}(\hat{S}(\hat{t}_p))}{[\hat{f}(\hat{t}_p)]^2} \quad (\text{eq. 21})$$

Le numérateur est l'estimateur de la variance de la survie estimée de Greenwood. Le dénominateur est l'estimateur de la fonction de densité.

Ainsi l'intervalle de confiance d'un quantile estimé est obtenu par :

$$\hat{t}_p \pm z_{1-\alpha/2} \hat{\sigma}(\hat{t}_p) \quad (\text{eq. 22})$$

#### 5. La survie médiane résiduelle estimée

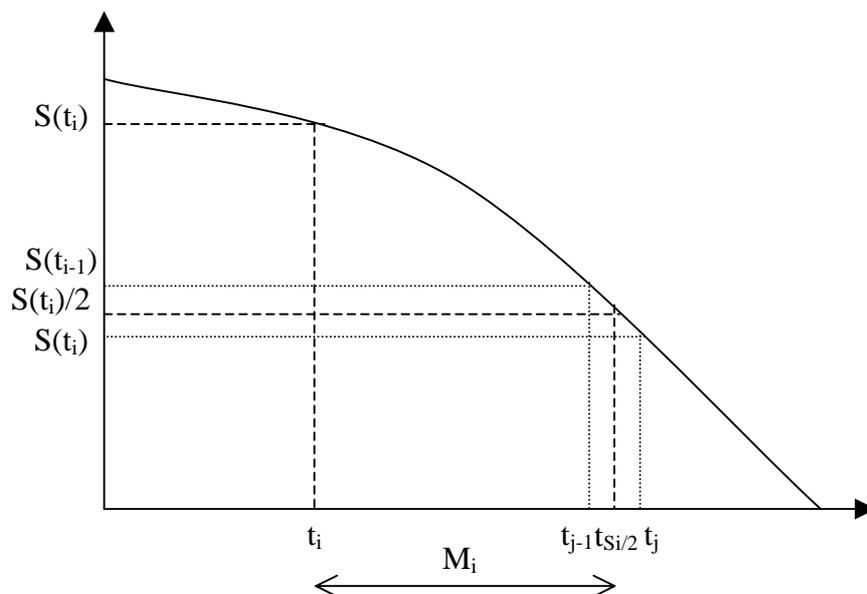
Un indicateur particulier de l'analyse des durées est la médiane, préféré à la moyenne en raison de la censure de certaines observations. L'estimation non-paramétrique produit pour la population à risque à chaque date le temps de survie supplémentaire atteint par 50% de ces individus. Cette survie médiane résiduelle à la date  $t$  exprime donc le temps de survie médian restant pour la population à risque à cette date.

Elle est donnée par :

$$\hat{M}_i = t_{j-1} - t_i + b_j \frac{\hat{S}(t_{j-1}) - \hat{S}(t_i)/2}{\hat{S}(t_{j-1}) - \hat{S}(t_j)} \quad (eq. 23)$$

où  $[t_{j-1}, t_j[$  est l'intervalle défini par :  $\hat{S}(t_{j-1}) \geq \hat{S}(t_i)/2 > \hat{S}(t_j)$ . Ce qui correspond à une application du théorème de Thalès dans la figure 2. A la date  $t_i$ , la population à risque représente une proportion de l'échantillon initial donnée par  $S(t_i)$ . Le temps de survie médian de cette population à risque en  $t_i$  correspond donc à  $S(t_i)/2$ . L'approximation de ce temps médian est obtenu par interpolation linéaire entre les deux dates observées les plus proches encadrant le temps de survie. Enfin, le temps de survie médian résiduel est la différence ( $M_i$ ) entre le temps de survie médian de la population à risque à la date considérée ( $t_{Si/2}$ ) et la date déjà atteinte par ces individus ( $t_i$ ).

**figure 2 : Détermination de la survie médiane résiduelle**



#### 6. Tests d'équivalence des survies selon les classes

Cette approche produit une approximation purement empirique des fonctions de hasard et de survie qui est donc *a priori* proche de la réalité. Mais, elle modélise difficilement l'effet des covariables. Dans notre application, seuls des tests ont pu être mis en œuvre pour examiner les relations entre les durées et les covariables.

Les tests d'équivalence des survies entre classes sont basés sur les tableaux de contingence à chaque date  $t_i$ , à partir desquels sont testées les différences entre le nombre

d'interruptions observées dans une classe  $j$  :  $d_j(t_i)$  et le nombre d'interruptions prédites  $\hat{e}_j(t_i)$  à partir de l'estimation d'une survie commune à toutes les classes.

Le tableau de contingence dans le cas d'une variable dichotomique à la date  $t_i$  est :

Événement	Classe 1	Classe 0	Total
<b>Interruption</b>	$d_1(t_i)$	$d_0(t_i)$	$d(t_i)$
<b>Non interruption</b>	$n_1(t_i)-d_1(t_i)$	$n_0(t_i)-d_0(t_i)$	$n(t_i)-d(t_i)$
<b>Population à risque</b>	$n_1(t_i)$	$n_0(t_i)$	$n(t_i)$

L'estimation du nombre d'interruptions dans la classe 1 à chaque date  $t_i$  est :

$$\hat{e}_1(t_i) = \frac{n_1(t_i)d(t_i)}{n(t_i)} \quad (eq. 24)$$

Et la variance de  $d_1(t_i)$  est usuellement estimée par (Hosmer et Lemeshow, 1999) :

$$v_1(t_i) = \frac{n_1(t_i) n_0(t_i) d(t_i) (n(t_i) - d(t_i))}{n^2(t_i) (n(t_i) - 1)} \quad (eq. 25)$$

Ainsi, la statistique de test suivante suit un chi-deux à 1 degré de liberté :

$$Q = \frac{\left[ \sum_{i=1}^m w_i (d_1(t_i) - \hat{e}_1(t_i)) \right]^2}{\sum_{i=1}^m w_i^2 \hat{v}_1(t_i)} \quad (eq. 26)$$

où  $t_m$  est la dernière date observée.

Selon les pondérations choisies  $w_i$ , plusieurs versions du test peuvent être déduites. Les pondérations sont le plus souvent fixées à 1. Il s'agit alors du test du *log-rank*. Le test de Wilcoxon (ou Wilcoxon généralisé) correspond à des pondérations égales à la population à risque à chaque date. Le choix de ces pondérations influence le résultat du test, dans la mesure où elles privilégient certaines différences entre les survies selon la proximité avec le début ou la fin du processus. Ainsi, les pondérations du test *log-rank* privilégient les différences de survies apparaissant au début du processus. Et le test de Wilcoxon privilégie les différences de survies des dates proches de la fin du processus.

### III. Les modèles de régression

Dans l'objectif de modéliser la distribution de la durée de vie en fonction de variables exogènes deux types de modèles de régression paramétriques peuvent être utilisés : le modèle

à durée de vie accélérée ou le modèle à hasards proportionnels. Ces deux modèles se distinguent par les hypothèses qu'ils formulent sur la spécification de la relation entre la durée de vie et les covariables. Le premier suppose un effet multiplicatif direct sur la durée. Le second modèle suppose un effet multiplicatif des covariables sur le hasard.

Comme nous allons le voir, le modèle à durée de vie accélérée propose la forme la plus flexible, mais reste dépendant de la validité de la distribution choisie. Son estimation pleinement paramétrique sera réalisée par la technique du maximum de vraisemblance. Le modèle à hasards proportionnels impose une contrainte supplémentaire sur la forme du hasard estimé. Cependant, cette forme particulière permet l'estimation du modèle par la méthode semi-paramétrique de Cox basée sur la vraisemblance partielle de Cox, qui ne nécessite pas la spécification de la distribution des durées. Ce type de modèle sera aussi estimable par maximum de vraisemblance si la distribution supposée soutient l'hypothèse de proportionnalité des hasards.

#### 1. Approche paramétrique – Modèle à durée de vie accélérée - Modèle log-linéaire

Les modèles paramétriques supposent une structure log-linéaire du modèle. Le logarithme de la durée ( $T$ ) est modélisé par :

$$\ln t = g(X, \beta) + \sigma \varepsilon \quad (\text{eq. 27})$$

où  $X$  est la matrice composée des vecteurs colonne des covariables,  $\beta$  est le vecteur colonne des coefficients associés aux covariables et  $\varepsilon$  est un terme d'erreur dont nous supposons la distribution connue (par exemple, distribution normale, logistique ou distribution à valeurs extrêmes) et dont dépendront les fonctions décrivant la distribution de la durée  $T$ .

Selon le domaine d'application, la fonction  $g(X, \beta)$  est spécifiée différemment. Par exemple, les formes suivantes sont généralement utilisées en sciences physiques :

$$g(X, \mu \beta) = \mu + (\ln X) \beta \quad (\text{quand } X \text{ est un voltage})$$

$$g(X, \mu \beta) = \mu + (1/X) \beta \quad (\text{quand } X \text{ est une température})$$

La spécification la plus courante en économie est la suivante :

$$g(X, \mu \beta) = \mu + X\beta \quad (\text{eq. 28})$$

Elle permet une interprétation aisée des coefficients estimés. Si  $X=0$ , alors  $\mu$  représente l'indicateur de localisation de la variable aléatoire  $\ln(T)$  ( $E(g(X, \mu\beta)) = \mu + E(X\beta)$ ). Et les coefficients donne les variations de  $E[\ln T / X]$ .

La forme linéaire  $(X\beta)$  permet une estimation aisée des coefficients par maximum de vraisemblance. Et la transformation logarithmique assure la positivité des valeurs prédites. La durée,  $T$ , s'exprime alors comme la transformation suivante de  $\varepsilon$ , qui autorise une grande liberté pour la relation entre  $T$  et les covariables :  $t = \exp\{g(X, \beta)\} \cdot (\exp\{\varepsilon\})^\sigma$

Par construction, ce modèle considère que les covariables affectent directement le temps de survie :

$$\sigma\varepsilon = \ln(t \cdot \exp\{-g(X, \mu\beta)\}) \quad (\text{eq. 29})$$

En posant  $\lambda = \exp\{-g(X, \mu\beta)\} = \exp\{-X\beta\}$ <sup>97</sup> (Cox et Oakes, 1988),  $\lambda$  a un rôle de facteur d'échelle. Si  $\lambda > 1$ , l'échelle de temps est accélérée. Et elle est ralentie si  $\lambda < 1$ . Les covariables expliquent une survie plus ou moins longue<sup>98</sup>.

Dans ce modèle la fonction de survie est de la forme :

$$S(t|X) = S_0(t \exp\{-X\beta\}) \quad (\text{eq. 30})$$

où  $S_0(t) = S(t|X=0)$  est la fonction de survie de référence qui est donnée par la distribution supposée de la durée. De plus, la fonction de hasard correspondante est donnée par :

$$\begin{aligned} h(t|X) &= \frac{-\partial S(t|X)/\partial t}{S(t|X)} = -\frac{d \ln S(t)}{dt} \quad (\text{eq. 31}) \\ &= h_0(t \exp\{-X\beta\}) \cdot \exp\{-X\beta\} \end{aligned}$$

Le hasard de référence est fonction de l'interaction entre la durée et les variables exogènes.

L'interprétation des coefficients estimés peut se faire par le ratio des durées pour les variables dichotomiques (par exemple, pour une covariable binaire  $X_I$ ) :

---

<sup>97</sup> Dans cette expression, le signe négatif du vecteur des coefficients découle de la construction du modèle. De plus, il permet l'interprétation directe des coefficients en termes d'effet sur la durée. Pour son interprétation en termes d'effet sur le hasard, l'opposé du coefficient doit être considéré. Par exemple, un coefficient positif aura un effet négatif sur le hasard. Il réduit donc la probabilité d'interruption et par conséquent, allonge la durée de survie.

<sup>98</sup> Allison (1995, p.62) utilise l'image du rapport entre l'âge d'un chien et l'âge d'un homme qui vaut  $\lambda=7 : 1$  an humain pour 7 ans canins.

$$\frac{E[T|X_1 = 1]}{E[T|X_1 = 0]} = \frac{E[\exp\{\mu + \beta_1 1 + \sigma\varepsilon\}]}{E[\exp\{\mu + \sigma\varepsilon\}]} = \exp\{\beta_1\} \quad (eq. 32)$$

ou en termes relatifs pour les variables continues car :

$$\frac{\partial \ln t}{\partial X_k} = \beta_k \quad (eq. 33)$$

## 2. Modèle de Cox – Approche semi-paramétrique - Modèle à hasards proportionnels

Nous avons vu que les modèles paramétriques imposent une distribution *a priori* aux durées. Le modèle à hasards proportionnels ajoute une décomposition de la fonction de hasard de la forme suivante :

$$h(t/X) = h_0(t) \cdot g(X, \beta) \quad (eq. 34)$$

où  $h_0(t)$  est la fonction de hasard de référence. Et  $g(X)$  est la fonction des effets des covariables.  $h_0(t)$  et  $g(X, \beta)$  sont telles que le taux de hasard,  $h(t)$  est positif. Et  $h(t) = h_0(t)$  quand  $g(X, \beta) = 1$  et  $g(X=0, \beta) = 1$ .

$h_0(t)$  est uniquement fonction du temps de survie et représente la variation de la probabilité conditionnelle d'interruption dans le temps indépendamment des covariables. La forme de cette fonction de référence sera déduite de la fonction de distribution supposée des durées. Comme nous le verrons, cette propriété de décomposition des effets du temps et des covariables n'est pas soutenue par l'ensemble des distributions. Seules les distributions exponentielle et Weibull supportent le modèle à hasards proportionnels, les autres distributions conduisant à un modèle à durée de vie accélérée.

$g(X, \beta)$  modifie le hasard de référence,  $h_0(t)$ , en fonction des covariables. Généralement cette fonction des covariables est définie comme :  $g(X, \beta) = \exp(X\beta)$  (Cox, 1972)<sup>99</sup>.

Le modèle estimé est :

$$\ln h(t) = \ln h_0(t) + X\beta \quad (eq. 35)$$

---

<sup>99</sup> L'estimation dans le modèle de Cox se concentre sur le hasard. De ce fait, le signe positif du vecteur de coefficients permet une interprétation en termes d'effets des covariables sur le hasard. L'opposé du coefficient donne donc l'impact sur la durée.

La séparation de l'effet du temps et des effets des covariables conduit le modèle à hasards proportionnels à supposer la proportionnalité des taux de hasards de deux individus,  $i$  et  $j$ , ayant des attributs différents. Etant donné que les effets des covariables ne sont pas dépendants du temps, le rapport des hasards de deux individus,  $i$  et  $j$  est :

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t) \cdot g(X_i, \beta)}{h_0(t) \cdot g(X_j, \beta)} = \frac{h_0(t) \cdot \exp\{X_i \beta\}}{h_0(t) \cdot \exp\{X_j \beta\}} = \exp\{(X_i - X_j) \beta\} \quad (eq. 36)$$

L'estimation directe du hasard produit d'une part, les paramètres de la fonction de distribution des durées et donc ceux du hasard de référence  $h_0(t)$ , et d'autre part, les coefficients  $\beta$  associés aux covariables<sup>100</sup>. Cependant, l'élimination du hasard de référence dans l'expression du quotient de hasards offre aussi la possibilité d'estimer les coefficients associés aux covariables par la vraisemblance partielle de Cox. L'expression de cette dernière est un produit de quotients de hasards et ne nécessite donc pas l'estimation des attributs d'une distribution supposée.

Ces coefficients estimés peuvent s'interpréter, soit en termes d'effet sur le ratio des hasards, soit par l'effet de la covariable sur le logarithme du hasard, donc en variation relative du hasard :

$$\frac{\partial \ln h(t)}{\partial X_k} = \frac{\partial \ln [h_0(t) \cdot g(X, \beta)]}{\partial X_k} = \frac{\partial \ln g(X, \beta)}{\partial X_k} = \frac{\partial \ln (\exp\{X \beta\})}{\partial X_k} = \beta_k \quad (eq. 37)$$

Ainsi, pour une variable dichotomique, un coefficient estimé  $\beta = \ln(2)$  s'interprètera de préférence en termes de rapport de hasards. Le hasard évalué pour une modalité de la covariable est deux fois plus grand que le hasard pour l'autre modalité. Pour une variable quantitative, un coefficient positif implique alors qu'une augmentation de la covariable associée augmente le taux de hasard<sup>101</sup> et diminue en conséquence le temps de survie estimé. Ainsi, si la covariable  $j$  augmente de 1 unité, alors le hasard varie de  $100(e^{\beta_j} - 1)\%$ .

---

<sup>100</sup> L'élimination du hasard de référence dans l'expression du quotient de hasards offre la possibilité d'estimer les coefficients associés aux covariables par la vraisemblance partielle de Cox. L'expression de cette dernière est un produit de quotients de hasards et ne nécessite donc pas l'estimation des attributs d'une distribution supposée.

<sup>101</sup> Le signe moins devant le  $\beta$  dans la fonction  $g_0(\cdot)$ .

### Test de l'hypothèse de proportionnalité des hasards

Le modèle de Cox repose sur l'hypothèse de proportionnalité des hasards. Il est donc nécessaire de tester la validité de cette hypothèse.

La structure du modèle à hasards proportionnels suppose que pour deux individus 1 et 2 caractérisés par  $X_1$  et  $X_2$ , la relation entre leurs hasards peut être définie par :

$$h_1(t/X_1) = \delta h_2(t/X_2) \quad (\text{eq. 38})$$

où  $\delta$  est une constante de proportionnalité. Nous avons vu que le taux de proportionnalité ( $\delta$ ) de ces taux de hasard est une fonction des différences de caractéristiques des individus (eq. 35).

Cette hypothèse de proportionnalité conduit à des survies liées par :

$$S_1(t/X_1) = [S_2(t/X_2)]^\delta \quad (\text{eq. 39})$$

En conséquence, la proportionnalité des hasards peut être testée graphiquement en représentant soit la transformation des survies suivante<sup>102</sup> :

$$\log[-\log S_1(t/X_1)] = \log(\delta) + \log[-\log S_2(t/X_2)] \quad (\text{eq. 40})$$

soit la transformation des hasards :

$$\log[h_1(t/X_1)] = \log(\delta) + \log[h_2(t/X_2)] \quad (\text{eq. 41})$$

Pour chaque transformation, les courbes des survies ou des hasards ne diffèrent que de la constante,  $\log \delta$ . Leurs représentations graphiques face à  $\ln t$  seront alors des droites parallèles, si les hasards estimés vérifient l'hypothèse de proportionnalité

---

<sup>102</sup> Il s'agit aussi d'une transformation du hasard intégré, qui se définit par  $H(t/X) = -\log S(t/X)$

figure 3 : Illustration de hasards proportionnels

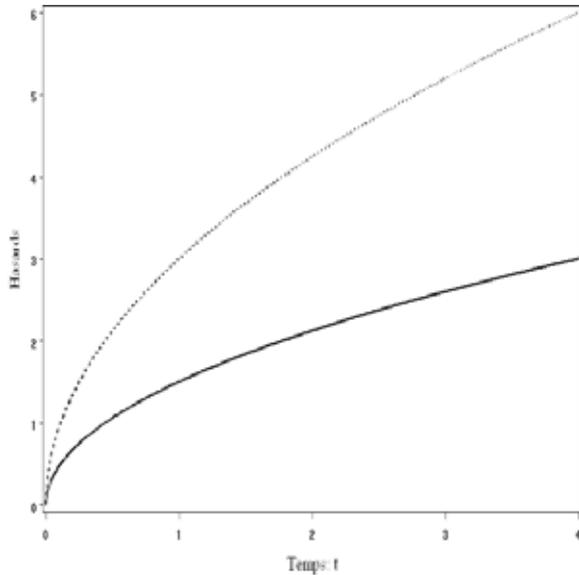
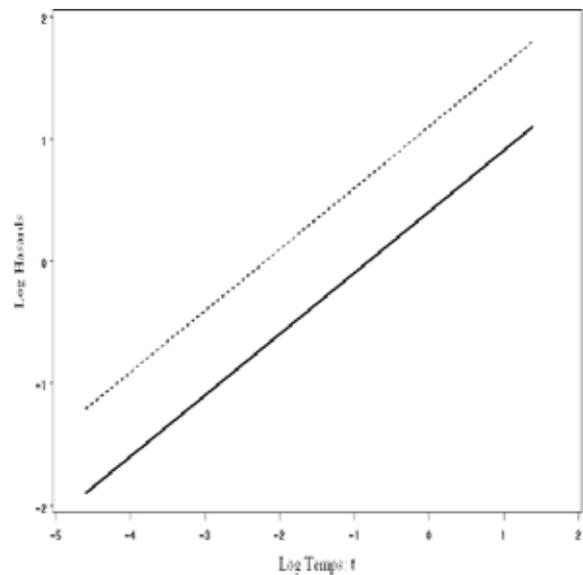


figure 4 : Test de proportionnalité par les logarithmes des hasards



### 3. Estimations par maximum de vraisemblance et maximum de vraisemblance partielle

L'estimation de ces modèles paramétriques est réalisable par la méthode du maximum de vraisemblance. Nous laissons au lecteur le soin de consulter les ouvrages de référence sur les principes et les propriétés du maximum de vraisemblance<sup>103</sup>. Dans un premier temps, nous présentons ici les fonctions de vraisemblance et de log-vraisemblance estimées dans le cadre des modèles à durée de vie accélérée, dont les expressions en présence d'observations censurées divergent de la forme classique. Ensuite, nous présentons l'estimation semi-paramétrique par le maximum de la vraisemblance partielle de Cox.

#### a) Estimation paramétrique

L'approche paramétrique suppose une distribution *a priori* des résidus du modèle :  $\ln t = X\beta + \sigma\varepsilon$ , où  $\varepsilon$  est le terme d'erreur. De la même façon que dans un modèle linéaire standard, les résidus sont supposés suivre une loi de distribution connue de moyenne et de

<sup>103</sup> Greene (1997), chap. 4.

variance constante entre observations. Nous supposons connaître la fonction de distribution  $F(\varepsilon)$ , la densité  $f(\varepsilon)$  et la fonction de survie  $S(\varepsilon)$  caractérisant la distribution de  $\varepsilon$ . La présence du paramètre  $\sigma$  simplifie l'estimation du modèle car il permet de fixer la variance des résidus à une valeur standard,  $\sigma$  étant adapté afin de compenser les variations des variances d'un échantillon à un autre.

A partir de la distribution supposée des résidus et par changement de variable nous pouvons déduire l'expression de la densité de la durée :

$$\begin{aligned} f_t(t) &\equiv f_\varepsilon(\varepsilon = g(t)) \cdot \frac{\partial g(t)}{\partial t} \\ &= \frac{1}{\sigma} \frac{1}{t} f_\varepsilon\left(\frac{\ln t - X\beta}{\sigma}\right) \end{aligned} \quad (eq. 42)$$

$$\text{où } \varepsilon = g(t) = \frac{\ln t - X\beta}{\sigma}$$

Ainsi, si toutes les durées  $t_i$  sont observées, la vraisemblance de l'échantillon s'écrit :

$$L = \prod_{i=1}^N f_t(t_i) \quad (eq. 43)$$

Dans le cas de censure à droite, en indexant par  $j=1, \dots, J$  les observations complètes et par  $k=1, \dots, K$  les observations censurées à droite, la vraisemblance est donnée par :

$$L = \prod_{j=1}^J f_t(t_j) \prod_{k=1}^K S_t(t_k) \quad (eq. 44)$$

Soit en logarithme :

$$\begin{aligned} \ln L &= \sum_{j=1}^J \ln f_t(t_j) + \sum_{k=1}^K \ln S_t(t_k) \\ &= \sum_{j=1}^J \ln(h_t(t_j) S_t(t_j)) + \sum_{k=1}^K \ln S_t(t_k) \\ &= \sum_{j=1}^J \ln h_t(t_j) + \sum_{i=1}^N \ln S_t(t_i) \\ &= \sum_{i=1}^N [\delta_i \ln h_t(t_i) + \ln S_t(t_i)] \end{aligned} \quad (eq. 45)$$

où

$$\delta_i = \begin{cases} 1 & \text{si l'observation est complète} \\ 0 & \text{si l'observation est censurée à droite} \end{cases}$$

Comme nous l'avons déjà évoqué, selon le type de distribution supposée pour la distribution de référence des durées, le modèle estimé pourra être un modèle à durée de vie accéléré ou un modèle à hasards proportionnels.

b) Estimation semi-paramétrique du modèle à hasards proportionnels

Le modèle dit de Cox permet de réaliser l'estimation d'un modèle à hasards proportionnels, sans spécifier la distribution des durées. Contrairement aux approches plus classiques utilisées dans la régression ordinaire ou les modèles à durée de vie accélérée, cette méthode ne modélise pas l'effet direct des covariables sur la valeur de la variable aléatoire, mais leur effet sur la valeur de la fonction de hasard.

En notant les événements observés par ordre d'apparition,  $t_1 < \dots < t_K$  et  $X_{(k)}$  le vecteur de variables exogènes de l'individu dont le temps d'interruption est égal à  $t_k$ , et  $R_{(ti)}$ , l'ensemble des individus à risque à  $t_i$ , la vraisemblance partielle de Cox est donnée par :

$$PL(\beta) = \prod_{k=1}^K \frac{h_{(k)}}{\sum_{i \in R(t_k)} h_{(i)}} = \prod_{k=1}^K \frac{\exp\{X_{(k)}\beta\}}{\sum_{i \in R(t_k)} \exp\{X_{(i)}\beta\}} \quad (eq. 46)$$

L'expression de la vraisemblance partielle diffère de la vraisemblance classique par l'index du produit et la probabilité considérée. Ici le produit est réalisé sur les événements et non sur les individus. Et la probabilité considérée est celle qu'un individu donné subisse l'événement  $k$  considéré, sachant qu'il fait partie de la population à risque à cette date<sup>104</sup>. Cette probabilité est donnée par le ratio des hasards.

Pour les détails de la maximisation de cette vraisemblance et l'introduction de la censure, le lecteur pourra se référer à Kalbfleisch et Prentice (1980).

L'estimation semi-paramétrique est basée sur une fonction de vraisemblance partielle qui ne dépend que des paramètres associés aux variables exogènes. Ainsi l'approche semi-paramétrique complète l'approche non-paramétrique, qui ne permet d'analyser que la distribution des durées et révèle donc uniquement des informations relatives aux formes des fonctions de hasard et de survie.

---

<sup>104</sup> Voir Collet (1994) pour plus de détails

Le modèle de Cox offre une alternative satisfaisante à l'estimation paramétrique pour l'estimation des effets des variables exogènes sans prendre de risque sur la distribution des durées. L'usage de la fonction de vraisemblance partielle proposée par Cox (1972) évite donc les risques d'une mauvaise spécification. Dans ce cas, la qualité de l'estimation des effets des covariables est considérée plus robuste dans cette approche que dans l'approche pleinement paramétrique dont la distribution supposée est erronée (Oakes, 1977). Par contre, ce type d'estimation produit théoriquement des estimateurs dont les variances sont plus élevées que ceux produits par un modèle pleinement paramétrique dont la distribution des durées est connue. Cependant, plusieurs études ont montré que cette la perte d'efficacité des estimateurs reste réduite (Hensher et Mannering, 1994). Efron (1997) et Oakes (1977) ont obtenu des matrices de variance-covariance des estimateurs du modèle de Cox proches de ceux du modèle pleinement paramétrique.

En définitive, le modèle de Cox peut être appliqué sans perte dommageable d'efficacité des estimateurs. Mais, il exclut l'estimation directe de la fonction de hasard de référence et ne permet pas de considérer la dynamique du processus. De plus, cette estimation semi-paramétrique repose sur l'hypothèse forte de proportionnalité des hasards.

Cette approche peut être complétée par l'estimation non-paramétrique pour obtenir un estimateur de la fonction de survie. Dans le modèle à hasards proportionnels, la survie en  $t$  pour un individu caractérisé par  $X$  peut s'écrire :

$$S(t)=[S_0(t)]^{\exp\{X\beta\}} \quad (eq. 47)$$

où  $S_0(t)$  est la fonction de survie de référence, qui est la valeur de la survie pour un individu dont les covariables sont toutes nulles. Après avoir estimé les coefficients  $\beta$  par maximisation de la vraisemblance partielle, un estimateur de  $S_0(t)$  peut être obtenu par la méthode non-paramétrique. Ainsi un estimateur de la survie peut être reconstitué. Par convention, la survie estimée est évaluée au point moyen des covariables.

A l'aide de cette estimation « mixte » nous pouvons réaliser les tests relatifs à la forme du hasard et les analyses des résidus, tous deux présentés dans la partie suivante consacrée à l'approfondissement de l'estimation paramétrique.

4. Estimation paramétrique - modèles à durée de vie accélérée et modèles à hasards proportionnels

Nous avons vu que les modèles paramétriques reposent sur l'hypothèse d'une distribution des durées. Cependant, ils offrent une grande variété de formes pour les fonctions de hasard et de survie. Selon la distribution supposée, ils permettent de considérer que les covariables agissent directement sur la durée du processus ou sur le hasard de référence. Enfin, la spécification complète des fonctions de hasard et de survie permettra de produire des prévisions.

a) Présentation des distributions usuelles

Les distributions usuelles pour les durées sont présentées dans le tableau 1. Chacune est déterminée à partir de la distribution supposée du terme d'erreur du modèle log-linéaire de la durée. Le tableau 1 présente les correspondances entre la distribution des résidus et la distribution des durées<sup>105</sup>.

**tableau 1 : Distributions des résidus et distributions des durées**

Distribution des résidus ( $\varepsilon$ )	Distribution des durées (T)	Type de modèle
Valeurs extrêmes à 1 paramètre	Exponentielle	HP et durée de vie accélérée
Valeurs extrêmes à 2 paramètres	Weibull	HP et durée de vie accélérée
Logistique	Log-logistique	Durée de vie accélérée
Normale	Log-normale	Durée de vie accélérée
Log Gamma à 3 paramètres	Gamma généralisée	Durée de vie accélérée

HP : hasards proportionnels

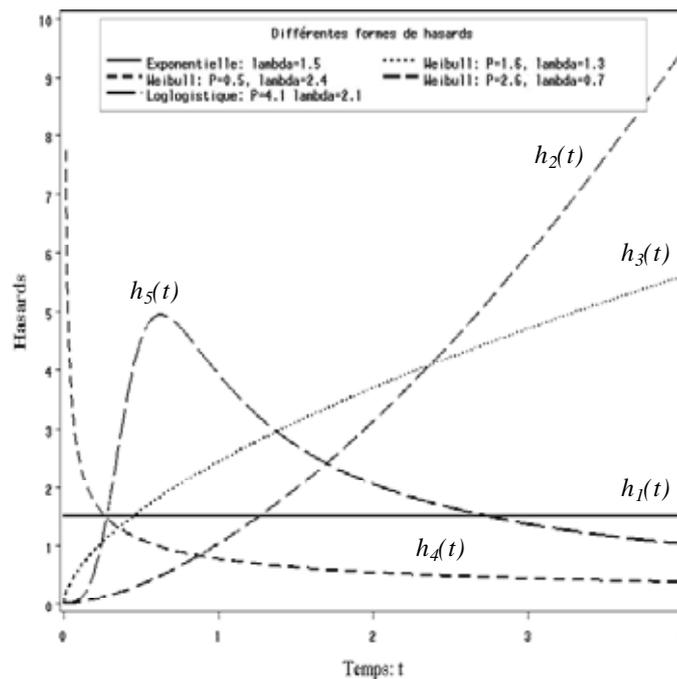
Ces différentes distributions permettent l'ajustement de nombreuses formes de hasards. Par exemple à partir des hasards les plus simples, comme ceux des modèles exponentiel, Weibull et log-logistique, le hasard peut être constant, croissant, décroissant, concave, convexe et non-monotone (figure 5). Le hasard exponentiel ( $h_1(t)$ ) est constant et caractérise donc les processus indépendants du temps. Le hasard Weibull ( $h_2(t)$ ,  $h_3(t)$ ,  $h_4(t)$ ) est monotone. S'il est positif, alors plus longtemps un individu reste dans un état, plus la probabilité d'en sortir est grande. Enfin, le hasard log-logistique ( $h_5(t)$ ) permet les formes monotones et non-monotones

---

<sup>105</sup> Allison (1995), p. 66.

selon variance de la variable aléatoire caractérisée. La dynamique temporelle décrite révèle deux phases. Dans un premier temps, la probabilité de sortie est croissante. Ensuite, cette probabilité décroît, l'individu voit sa probabilité de sortie se réduire avec le temps écoulé.

**figure 5 : Exemples de formes de hasard correspondant aux distributions : exponentielle ( $h_1(t)$ ), Weibull ( $h_2(t)$ ,  $h_3(t)$ ,  $h_4(t)$ ), log-logistique ( $h_5(t)$ ).**



Le hasard de la distribution exponentielle est constant. La distribution Weibull est une généralisation de la distribution exponentielle. Elle est plus flexible pour capturer la dynamique temporelle, mais reste limitée aux formes monotones. Enfin, la distribution log-logistique permet des formes de hasards non-monotones et l'interprétation de ces coefficients est relativement simple. De ce fait elle est souvent utilisée comme une approximation de la distribution log-normale ou gamma généralisée.

Ainsi, connaissant la fonction de densité de  $\varepsilon$  et par changement de variable, nous pouvons déduire la fonction de densité caractérisant la distribution de  $T$ .

Pour la présentation des différents modèles, nous posons :  $\lambda = \exp\{-g(X, \beta)\}$ ,  $g(X, \beta) = X\beta$  et  $p = 1/\sigma$ .

### Modèle Exponentiel

Le modèle exponentiel impose la contrainte  $\sigma = 1$ . Le terme résiduel  $\varepsilon$  suit une distribution à valeurs extrêmes standard (ou distribution de Gumbel ou encore distribution à double exponentielle) caractérisée par la densité :  $f_\varepsilon(\varepsilon) = \exp\{\varepsilon - \exp(\varepsilon)\}$  et une survie :  $S_0(\varepsilon) = \exp\{-\exp(\varepsilon)\}$ . Ce qui signifie que  $\ln T$  sachant  $X$  suit la même distribution. Et  $T$  a une distribution exponentielle de paramètre  $\lambda$ .

Le modèle exponentiel est caractérisé par les fonctions suivantes.

La fonction de densité exponentielle :

$$f_t(t) = \lambda \cdot \exp\{-\lambda t\} \quad (eq. 48)$$

La fonction de répartition :

$$F(t) = 1 - \exp\{-\lambda t\} \quad (eq. 49)$$

La fonction de survie :

$$S(t) = \exp\{-\lambda t\} \quad (eq. 50)$$

La fonction de hasard :

$$h(t) = \lambda \quad (eq. 51)$$

La fonction de hasard intégré :

$$H(t) = \lambda t \quad (eq. 52)$$

Le hasard est constant et indépendant de la durée du processus. Ainsi, le modèle exponentiel supporte la forme d'un modèle à hasards proportionnels. Le coefficient de proportionnalité  $\delta$  de eq. 37 est donné par  $\delta = \lambda$ . De plus, par construction du modèle, les covariables ont un effet direct sur la durée. Le modèle exponentiel est aussi un modèle à durée de vie accélérée.

Les coefficients  $\beta$  estimés sont interprétables en termes d'effet sur le logarithme de la durée. Donc,  $\exp\{\beta\}$  est l'effet de la covariable associée sur la durée. Et  $\exp\{-\beta\}$  pourra s'interpréter directement en termes d'effet sur le hasard.

Modèle Weibull :

Le modèle Weibull relaxe la contrainte sur  $\sigma$  et suppose une distribution des résidus  $\varepsilon$  à valeurs extrêmes standard :  $S_0(\varepsilon) = \exp\{-\exp(\varepsilon)\}$ .

Ainsi, la distribution de  $T$  est une distribution de Weibull de paramètres de forme,  $p=1/\sigma$  et d'échelle  $\lambda$ .

Nous obtenons :

$$f_t(t) = \lambda \cdot p \cdot (\lambda t)^{p-1} \cdot \exp\{-(\lambda t)^p\} \quad (eq. 53)$$

$$F(t) = 1 - \exp\{-(\lambda t)^p\} \quad (eq. 54)$$

$$S(t) = \exp\{-(\lambda t)^p\} \quad (eq. 55)$$

$$h(t) = \lambda \cdot p \cdot (\lambda t)^{p-1} \quad (eq. 56)$$

$$H(t) = (\lambda t)^p \quad (eq. 57)$$

Le hasard Weibull est monotone et décroissant avec  $t$  si  $\sigma > 1$ . Et il est monotone croissant et convexe, si  $0 < \sigma < 0,5$  et monotone croissant concave si  $0,5 < \sigma < 1$ . Pour  $\sigma = 1$ , nous retrouvons le hasard exponentiel. Enfin, il correspond à une ligne droite passant par l'origine si  $\sigma = 0,5$ .

Le modèle Weibull est, par construction, un modèle à durée de vie accélérée. Et son hasard peut être décomposé en isolant les effets de  $t$  et les effets des covariables,  $\lambda$ . Par conséquent, c'est aussi un modèle à hasards proportionnels de coefficient de proportionnalité,  $\delta = \lambda^p$ .

Les coefficients  $\beta$  seront interprétables en termes d'effet sur le hasard après transformation par  $\exp\{-\beta\}$ . Et les tests de significativité seront valables pour les coefficients et leurs transformations. Enfin, la transformation du hasard par le logarithme montre que ce coefficient  $\sigma$  peut s'interpréter comme l'élasticité du hasard par rapport au temps.

### Modèle log-logistique

Le modèle log-logistique suppose que les résidus, et donc  $\ln T$  sachant  $X$ , suivent une distribution logistique :

$$S_0(\varepsilon) = \frac{1}{1 + \exp\{\varepsilon\}} \quad (\text{eq. 58})$$

Ainsi  $T$ , suit une distribution log-logistique caractérisée par :

$$f_t(t) = \lambda p \cdot (\lambda t)^{p-1} \cdot (1 + (\lambda t)^p)^{-2} \quad (\text{eq. 59})$$

$$F(t) = (1 + (\lambda t)^p)^{-1} - 1 \quad (\text{eq. 60})$$

$$S(t) = (1 + (\lambda t)^p)^{-1} \quad (\text{eq. 61})$$

$$h(t) = \frac{\lambda p \cdot (\lambda t)^{p-1}}{1 + (\lambda t)^p} \quad (\text{eq. 62})$$

$$H(t) = \ln(1 + (\lambda t)^p) \quad (\text{eq. 63})$$

Ce modèle permet un hasard non-monotone. Lorsque  $\sigma < 1$  le hasard passe par l'origine, croit jusqu'à son maximum, puis décroît vers 0. Le maximum est atteint pour  $t = [(p-1)^{1/p}] / \lambda$ . Quand  $\sigma > 1$ , le hasard est décroissant convexe, comme le hasard Weibull. Enfin, quand  $\sigma = 1$ , le hasard prend la valeur de  $\lambda$  en  $t=0$ , puis converge vers 0. Le ratio des hasards n'est pas constant, cette distribution ne supporte pas la forme des modèles à hasards proportionnels.

En plus des interprétations des coefficients propres aux modèles à durée de vie accélérée, la transformation de la fonction de survie suivante permet d'interpréter les coefficients en termes d'effet sur la côte de la survie<sup>106</sup> :

$$\begin{aligned} \ln \left[ \frac{S(t)}{1 - S(t)} \right] &= \ln \left( \left( [1 + (\lambda t)^p] \cdot \left[ 1 - \frac{1}{1 + (\lambda t)^p} \right] \right)^{-1} \right) \\ &= \ln((\lambda t)^{-p}) \\ &= -p \ln(\lambda t) \\ &= pX\beta - p \ln t \end{aligned} \quad (\text{eq. 64})$$

Nous retrouvons là, les interprétations classiques des coefficients des modèles logistiques.

---

<sup>106</sup> La côte d'un événement est définie comme le rapport des probabilités de l'événement et de son complémentaire.

b) Diagnostique graphique sur la forme du hasard à partir du hasard intégré

Les expressions des hasards intégrés pour les différentes distributions et certaines de leurs transformations permettent de « tester » graphiquement l'adéquation du type de modèle avec l'échantillon. En effet, elles montrent pour chaque modèle des propriétés particulières. Ainsi, l'estimation non-paramétrique de Kaplan-Meier va permettre la recherche de ces propriétés et donc l'identification du modèle paramétrique et de la distribution les plus adaptés à l'échantillon.

Le hasard intégré exponentielle est de la forme :

$$H(t) = \lambda t \quad (\text{eq. 65})$$

La représentation graphique du hasard intégré estimé face à  $t$  doit être une droite passant par l'origine, si le modèle exponentiel est adapté.

Le hasard intégré Weibull :

$$H(t) = (\lambda t)^p \quad (\text{eq. 66})$$

Donc son logarithme est :

$$\begin{aligned} \ln H(t) &= p \ln(\lambda t) \\ &= p \ln t - pX\beta \end{aligned} \quad (\text{eq. 67})$$

La représentation graphique de  $\ln(\hat{H}(t))$  face à  $\ln t$  valide le modèle Weibull si elle a l'allure d'une droite de pente  $p=1/\sigma$ .

Le hasard intégré log-logistique :

$$H(t) = \ln(1 + (\lambda t)^p) \quad (\text{eq. 68})$$

La transformation du hasard intégré suivante :

$$\begin{aligned} \ln(\exp\{H(t)\} - 1) &= p \ln(\lambda t) \\ &= p \ln t - pX\beta \end{aligned} \quad (\text{eq. 69})$$

conduit à une représentation graphique face à  $\ln t$  caractérisée par une droite de pente  $p=1/\sigma$ , si le modèle log-logistique est adapté à l'échantillon.

c) Qualité d'ajustement  
Résidus de Cox-Snell

Nous avons vu précédemment comment « tester » graphiquement l'adéquation entre le type de modèle appliqué et l'échantillon à partir de représentations de transformations du hasard intégré. Cependant, ces résultats reposent sur l'hypothèse d'une homogénéité des individus, impliquant l'indépendance des covariables par rapport au temps. Une solution pour intégrer l'effet des covariables est d'utiliser les indicateurs traditionnels de la qualité d'ajustement : les résidus.

Il existe de nombreuses définitions de résidus pour les modèles de durées (Collet, 1994). En effet, selon la méthode d'estimation utilisée, les modèles ne produisent pas une prédiction de l'espérance de la durée. C'est notamment le cas lors de l'application de la vraisemblance partielle de Cox.

Les résidus retenus habituellement sont ceux de Cox-Snell. Ils sont définis par :

$$e_i = -\ln \hat{S}(t_i | X_i) \equiv \hat{H}(t_i | X_i) \quad (eq. 70)$$

où  $t_i$  est la date d'interruption ou la date censurée pour l'individu  $i$ . Et  $X_i$  est le vecteur des covariables de l'individu  $i$ . Cox et Snell (1968) ont montré que si la qualité de l'ajustement est acceptable, alors ces résidus sont approximativement distribués selon une distribution exponentielle de paramètre 1.

Test du rapport de vraisemblances

La comparaison des ajustements des différents types de modèles peut être réalisée par le test classique du rapport de vraisemblances. Cependant, ce test ne peut être réalisé que pour comparer des modèles emboîtés. Ce test ne permet donc pas de comparer l'ajustement du modèle log-logistique avec les autres modèles. En effet, seuls les modèles exponentiels, Weibull et log-normal sont des sous modèles dérivés du modèle gamma généralisé<sup>107</sup>.

La statistique du rapport de vraisemblances est :

$$LR = 2 \left[ \log L(\hat{\theta}_{H1}) - \log L(\hat{\theta}_{H0}) \right] \quad (eq. 71)$$

---

<sup>107</sup> Greene (1997), chap. 4, p. 161 ; Allison, (1995)

où  $\hat{\theta}_{H_0}$  et  $\hat{\theta}_{H_1}$  et sont les valeurs de l'ensemble des paramètres qui maximisent la fonction de log-vraisemblance sous les contraintes imposées par les hypothèses du test  $H_0$  et  $H_1$ , respectivement.

Sous  $H_0$ , cette statistique suit une loi de Khi-deux dont le nombre de degrés de liberté correspond au nombre de restrictions indépendantes dans  $H_0$ .

Les restrictions de la distribution gamma généralisée suivantes sont applicables<sup>108</sup> :

**tableau 2 : Distributions obtenues par restrictions de la distribution gamma généralisée**

Restrictions	Modèle / distribution	ddl
$\sigma=1$	Gamma standard	1
$\delta=1$ et $\sigma \neq 1$	Weibull	1
$\sigma=1$ et $\delta=1$	Exponentiel	2
$\delta \rightarrow 0$	Log-normal	1

d) Estimation des quantiles

Enfin, l'estimation paramétrique des durées permet de produire des prédictions. Parmi les indicateurs classiques, les quantiles peuvent être prédits selon la distribution estimée.

Le quantile  $q/100$  est défini par :

$$t_q = \{t | S(t) = q/100\} \quad (\text{eq. 72})$$

Modèles exponentiel et Weibull

Le modèle exponentiel étant une simplification du modèle Weibull, les quantiles pour le modèle exponentiel seront déduits des quantiles Weibull.

Ainsi le quantile d'ordre  $q$  dans le modèle Weibull est donnée par :

---

<sup>108</sup> La distribution gamma généralisée possède deux paramètres :  $\sigma$  et  $\delta$ . Sa densité est:

$$f(t) = \frac{p\lambda^{\frac{1}{\delta^2}} t^{p\frac{1}{\delta^2}-1} \exp\{-(\lambda t)^p\}}{\Gamma\left(\frac{1}{\delta^2}\right)}$$

$$\begin{aligned}
 t_q &= \frac{(-\ln S(t_q))^{\frac{1}{\sigma}}}{\lambda} & (eq. 73) \\
 &= (-\ln S(t_q))^{\sigma} \exp\{X\beta\}
 \end{aligned}$$

D'où la durée médiane :  $t_{50} = \frac{(\ln(2))^{\sigma}}{\lambda}$

Pour le modèle exponentiel, les quantiles sont obtenus en posant :  $\sigma=1$ .

L'effet des covariables sur ce temps de survie médian peut être déterminé. Pour une covariable binaire le ratio des survies médianes est donné par :

$$\frac{t_{50}(X_1 = 1)}{t_{50}(X_1 = 0)} = \exp\{\beta_1\} \quad (eq. 74)$$

Pour une variable continue, nous pouvons déduire une évaluation de l'élasticité de la médiane par rapport à la covariable  $X_k$ , de :

$$\begin{aligned}
 \frac{X_k}{t_{50}} \cdot \frac{\partial t_{50}}{\partial X_k} &= X_k \cdot \frac{\partial \ln(t_{50})}{\partial X_k} \\
 &= X_k \cdot \frac{\partial(\sigma \ln(\ln(2)) + X\beta)}{\partial X_k} & (eq. 75) \\
 &= X_k \beta_k
 \end{aligned}$$

Si  $X_k$  est le logarithme d'une variable  $Z_k$ , alors, le  $\beta_k$  associé correspond à l'élasticité de la médiane par rapport à cette variable  $Z_k$ . Cette élasticité est commune à tous les quantiles.

#### Modèle log-logistique

Pour le modèle log-logistique, les quantiles sont donnés par :

$$\begin{aligned}
 t_q &= \left(\frac{1-q}{q}\right)^{\sigma} \frac{1}{\lambda} \\
 &= \left(\frac{1-q}{q}\right)^{\sigma} \exp\{X\beta\} & (eq. 76)
 \end{aligned}$$

et la médiane par :

$$t_{50} = \frac{1}{\lambda} \quad (eq. 77)$$

Ainsi, pour une covariable binaire le ratio des survies médianes est donné par :

$$\frac{t_{50}(X_1 = 1)}{t_{50}(X_1 = 0)} = \exp\{\beta_1\} \quad (eq. 78)$$

L'effet des covariables sur la médiane est :

$$\frac{\partial t_{50}}{\partial X_k} = \beta \exp\{X\beta\} \quad (eq. 79)$$

Il s'interprète en termes d'élasticité de la médiane :

$$\frac{\partial \ln(t_{50})}{\partial \ln(X_k)} = \frac{X_k \beta_k}{\exp\{X\beta\}} \quad (eq. 80)$$

## Annexe X - Démonstrations des modèles exponentiel, Weibull, log-logistique

### I. Modèle Exponentiel

Le modèle exponentiel contraint le coefficient  $\sigma = 1$ . Le terme résiduel  $\varepsilon$  suit alors une fonction de distribution à valeurs extrêmes standard. Ce qui signifie que le log de  $t$  suit la même distribution. Et que  $T$  a une distribution exponentielle, d'où le nom du modèle. Ainsi, le modèle peut se réécrire :

$$\ln t = \beta_0 + \beta_1 X + \varepsilon^* \quad \Leftrightarrow \quad t = e^{\beta_0 + \beta_1 X} \varepsilon$$

avec  $\varepsilon^* = \ln \varepsilon$

$$f_w(w) = \exp\{w - \exp\{w\}\}$$

avec  $w = \ln t - X\beta$

Par changement de variable :

$$\begin{aligned} f_t(t) &= f_w(w = g(t)) \cdot \frac{\partial g(t)}{\partial t} \\ &= \exp\{\ln t - X\beta - \exp\{\ln t - X\beta\}\} \cdot \frac{1}{t} \\ &= \exp\{\ln t - X\beta\} \cdot \exp\{-\exp\{\ln t - X\beta\}\} \cdot \frac{1}{t} \\ &= (\exp\{\ln t\} \cdot \exp\{-X\beta\}) \cdot \exp\{-\exp\{(\ln t - X\beta)\}\} \cdot \frac{1}{t} \end{aligned}$$

en posant :  $\lambda = \exp\{-X\beta\}$

$$f_t(t) = \lambda \cdot \exp\{-\lambda t\}$$

Rque : fonction de densité exponentielle

$$\begin{aligned} F(t) &= \int_0^t f(u) du \\ &= \int_0^t \lambda \cdot \exp\{-\lambda u\} du \\ &= [-\exp\{-\lambda u\}]_0^t \\ &= 1 - \exp\{-\lambda t\} \end{aligned}$$

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= \exp\{-\lambda t\} \end{aligned}$$

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{\lambda \cdot \exp\{-\lambda t\}}{\exp\{-\lambda t\}} \\ &= \lambda \end{aligned}$$

$$\begin{aligned} H(t) &= -\ln(S(t)) \\ &= \lambda t \end{aligned}$$

La représentation graphique du hasard intégré face à  $t$  permet de vérifier l'adéquation de la distribution exponentielle avec l'échantillon.

## II. Modèle Weibull :

Le modèle Weibull suppose une distribution des résidus  $\varepsilon$  à valeurs extrêmes standard et relaxe la contrainte sur  $\sigma$ . Ainsi, la distribution de  $t$  est une distribution de Weibull.

Si  $\varepsilon \rightarrow EV$  alors la densité de la variable transformée :  $w = (\ln t - X\beta)/\sigma$  est de la forme :

$$f_w(w) = \exp\{w - \exp\{w\}\}$$

Par changement de variable :

$$\begin{aligned} f_t(t) &= f_w(w = g(t)) \cdot \frac{\partial g(t)}{\partial t} \\ &= \exp\left\{\frac{\ln t - X\beta}{\sigma} - \exp\left\{\frac{\ln t - X\beta}{\sigma}\right\}\right\} \cdot \frac{1}{\sigma} \cdot \frac{1}{t} \\ &= \exp\left\{\frac{\ln t - X\beta}{\sigma}\right\} \cdot \exp\left\{-\exp\left\{\frac{\ln t - X\beta}{\sigma}\right\}\right\} \cdot \frac{1}{\sigma} \cdot \frac{1}{t} \\ &= (\exp\{\ln t\} \cdot \exp\{-X\beta\})^{\frac{1}{\sigma}-1} \cdot \exp\{\ln t\} \cdot \exp\{-X\beta\} \cdot \exp\left\{-\exp\left\{(\ln t - X\beta)^{\frac{1}{\sigma}}\right\}\right\} \cdot \frac{1}{\sigma} \cdot \frac{1}{t} \end{aligned}$$

en posant :  $p = \frac{1}{\sigma}$  et  $\lambda = \exp\{-X\beta\}$

$$f_t(t) = \lambda \cdot p \cdot (\lambda t)^{p-1} \cdot \exp\{- (\lambda t)^p\}$$

Rque : fonction de densité Weibull

$$\begin{aligned} F(t) &= \int_0^t f(u) du \\ &= \int_0^t \lambda p (u \lambda)^{(p-1)} \cdot \exp\{- (\lambda u)^p\} du \\ &= \left[ - \exp\{- (\lambda u)^p\} \right]_0^t \\ &= 1 - \exp\{- (\lambda t)^p\} \end{aligned}$$

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= \exp\{- (\lambda t)^p\} \end{aligned}$$

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{\lambda \cdot p \cdot (\lambda t)^{p-1} \cdot \exp\{- (\lambda t)^p\}}{\exp\{- (\lambda t)^p\}} \\ &= \lambda \cdot p \cdot (\lambda t)^{p-1} \end{aligned}$$

$$\begin{aligned} H(t) &= -\ln(S(t)) \\ &= (\lambda t)^p \end{aligned}$$

D'où l'expression suivante :

$$\begin{aligned} \ln H(t) &= p \ln(\lambda t) \\ &= p \ln t - pX\beta \end{aligned}$$

dont la représentation graphique face à  $\ln t$  permet d'observer l'adéquation de la distribution Weibull avec l'échantillon.

### III. Modèle log-logistique

Le modèle log-logistique suppose que les résidus, et donc le log de  $t$ , suit une distribution logistique. Ainsi  $t$ , suit une distribution log-logistique. Ce modèle permet un hasard non-monotone.

Si  $\varepsilon \rightarrow \text{Logistique}$  alors la densité de la variable transformée : avec  $w = (\ln t - X\beta)/\sigma$ , est de la forme :

$$f_w(w) = \frac{\exp\{w\}}{1 + \exp\{w\}}$$

Par changement de variable :

$$\begin{aligned} f_t(t) &= \frac{\exp\left\{(\ln t - X\beta)\frac{1}{\sigma}\right\}}{t\sigma \cdot \left(1 + \exp\left\{(\ln t - X\beta)\frac{1}{\sigma}\right\}\right)^2} \\ &= \frac{p \cdot (\lambda t)^p}{t \cdot (1 + (\lambda t)^p)^2} \\ &= \lambda p \cdot (\lambda t)^{p-1} \cdot (1 + (\lambda t)^p)^{-2} \end{aligned}$$

$$\begin{aligned} F(t) &= \int_0^t \lambda p \cdot (u\lambda)^{p-1} \cdot (1 + (u\lambda)^p)^{-2} du \\ &= \left[ (1 + (u\lambda)^p)^{-1} \right]_0^t \\ &= (1 + (\lambda t)^p)^{-1} - 1 \end{aligned}$$

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= (1 + (\lambda t)^p)^{-1} \end{aligned}$$

$$\begin{aligned} h(t) &= \frac{\lambda p \cdot (\lambda t)^{p-1} \cdot (1 + (\lambda t)^p)^{-2}}{(1 + (\lambda t)^p)^{-1}} \\ &= \frac{\lambda p \cdot (\lambda t)^{p-1}}{1 + (\lambda t)^p} \end{aligned}$$

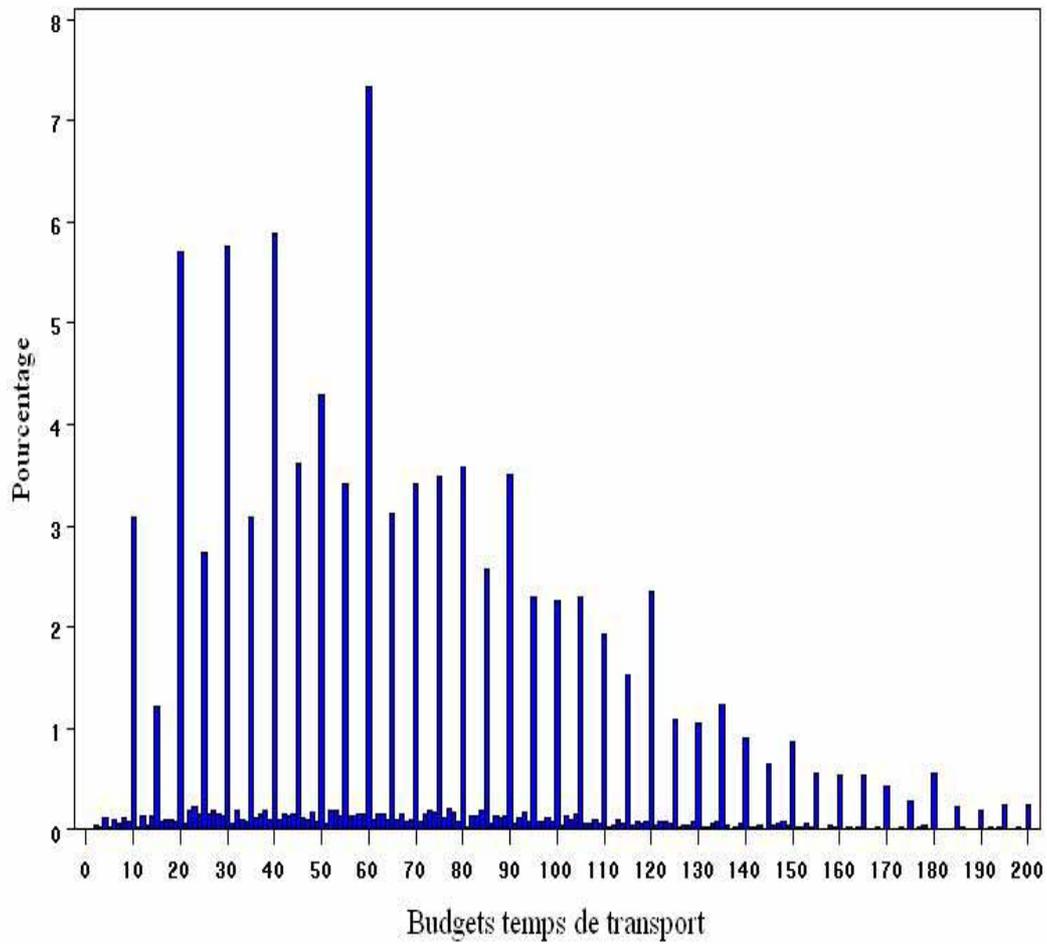
$$\begin{aligned} H(t) &= -\ln\left(\left(1 + (\lambda t)^p\right)^{-1}\right) \\ &= \ln\left(1 + (\lambda t)^p\right) \end{aligned}$$

D'où l'expression suivante :

$$\begin{aligned} \ln(\exp\{H(t) - 1\}) &= p \ln(\lambda t) \\ &= p \ln t - pX\beta \end{aligned}$$

dont la représentation graphique face à  $\ln t$  permet d'observer l'adéquation de la distribution log-logistique avec l'échantillon.

## Annexe XI - Présence d'arrondis dans les temps de transport déclarés



## Annexe XII - Régression MCO des budgets-temps de transport

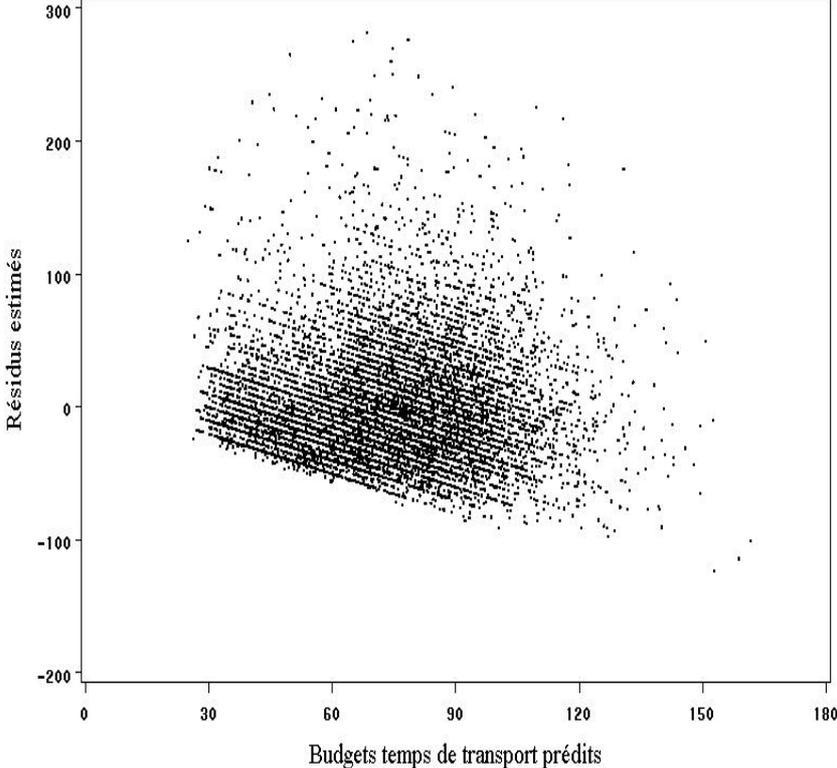
Analysis of Variance					
Source	DDL	Sum of Squares	Mean Square	Valeur F	Pr > F
<b>Model</b>	24	4611441	192143	106.68	<.0001
<b>Error</b>	11205	20181960	1801.16		
<b>Corrected Total</b>	11229	24793401			

Root MSE	42.44	R-Square	0.19
Dependent Mean	72.20	Adj R-Sq	0.18
Coeff Var	58.78		

Libellé	Parameter Estimate	Standard Error	t Value	Pr >  t	Standardized Estimate	Tolerance	Variance Inflation
Constante	62.006	2.354	26.34	<.0001	0	.	0
Nb d'enfants de plus de 5 ans	-1.962	0.643	-3.05	0.0023	-0.049	0.274	3.639
Nb d'enfants de moins de 5ans	-4.495	0.992	-4.53	<.0001	-0.045	0.706	1.416
Taille du ménage	2.710	0.479	5.66	<.0001	0.095	0.254	3.935
Homme	4.838	0.828	5.84	<.0001	0.051	0.935	1.068
Age	0.019	0.024	0.79	0.4275	0.008	0.648	1.542
Ménage à haut revenu	2.334	1.145	2.04	0.0416	0.018	0.882	1.133
Ménage à faible revenu	2.591	0.962	2.69	0.0071	0.025	0.831	1.202
Lundi	-2.211	1.052	-2.10	0.0356	-0.019	0.838	1.192
Vendredi	3.683	1.203	3.06	0.0022	0.028	0.859	1.163
Jeudi	2.868	1.096	2.62	0.0089	0.024	0.854	1.169
Centre-ville	1.821	1.506	1.21	0.2267	0.010	0.892	1.120
1ere couronne est	-2.151	1.178	-1.83	0.0679	-0.016	0.898	1.113
3eme couronne ouest	4.627	1.309	3.53	0.0004	0.031	0.904	1.105
3eme couronne est	-2.480	1.345	-1.84	0.0653	-0.016	0.905	1.104
Actif	14.103	1.394	10.11	<.0001	0.150	0.329	3.032
Deux-roues motorisées	-36.418	1.711	-21.28	<.0001	-0.299	0.366	2.732
Marche	-39.898	5.230	-7.63	<.0001	-0.067	0.933	1.071
Modes publics	4.983	1.752	2.84	0.0045	0.037	0.412	2.423
Vélo	-21.384	1.595	-13.40	<.0001	-0.219	0.271	3.683
Modes privés (auto)	-23.022	1.671	-13.77	<.0001	-0.199	0.344	2.900
Temps de travail	0.002	0.002	0.81	0.4197	0.012	0.316	3.160
Temps de loisir	0.091	0.004	20.88	<.0001	0.187	0.902	1.108
Temps d'achat	0.136	0.009	15.08	<.0001	0.140	0.834	1.197
Accompagnement(0/1)	17.437	1.226	14.22	<.0001	0.129	0.873	1.144

**Test of First and Second Moment  
Specification**

<b>DDL</b>	<b>Chi-Square</b>	<b>Pr &gt; ChiSq</b>
287	654.02	<.0001



## Annexe XIII - Test de normalité de la distribution des résidus

### Paramètres pour distribution Normal

Paramètre	Symbole	Estimation
Moyenne	Mu	0
Std Dev	Sigma	42.36182

### Tests d'ajustement pour distribution Normal

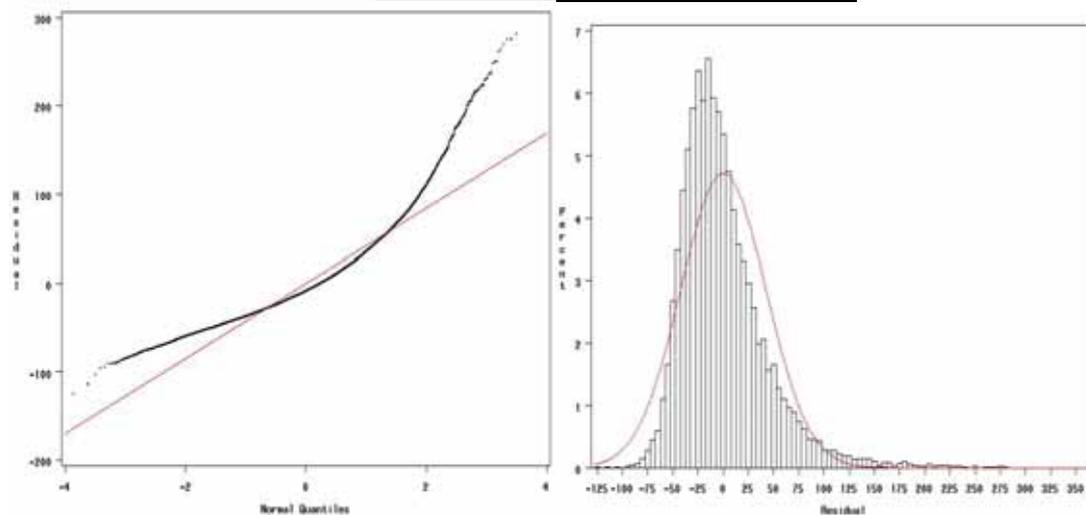
Test	Statistique	Seuil de significativité
<b>Kolmogorov-Smirnov</b>	D 0.093308	Pr > D <0.010
<b>Cramer-von Mises</b>	W-Sq 37.279503	Pr > W-Sq <0.005
<b>Anderson-Darling</b>	A-Sq 222.416037	Pr > A-Sq <0.005
<b>Jarque-Bera</b>	Chi 13647.56	Pr > D <0.001

Pour tester la normalité de la distribution des résidus, pour éviter les éventuels biais sur les grands échantillons des trois tests précédents automatiques sous SAS, la statistique de test de Jarque-Bera est calculée :

$$\left[ \frac{\text{skewness}^2}{6/n} + \frac{(\text{kurtosis} - 3)^2}{24/n} \right] \sim \chi^2(2)$$

### Quantiles pour distribution Normal

Pourcentage	Quantile	
	Observé	Estimé
1.0	-67.696	-98.548
5.0	-51.161	-69.678
10.0	-42.991	-54.288
25.0	-28.202	-28.572
50.0	-8.021	-0.000
75.0	19.074	28.572
90.0	52.225	54.288
95.0	77.379	69.678
99.0	147.993	98.548



## Annexe XIV - Test de normalité de la distribution des budgets-temps de transport

### Paramètres pour distribution Normal

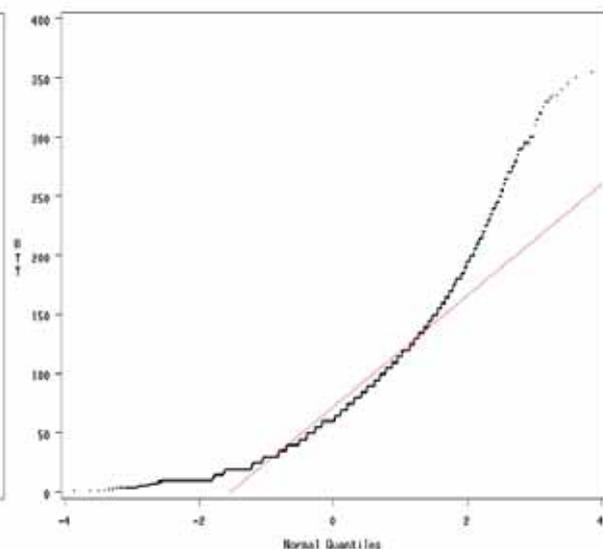
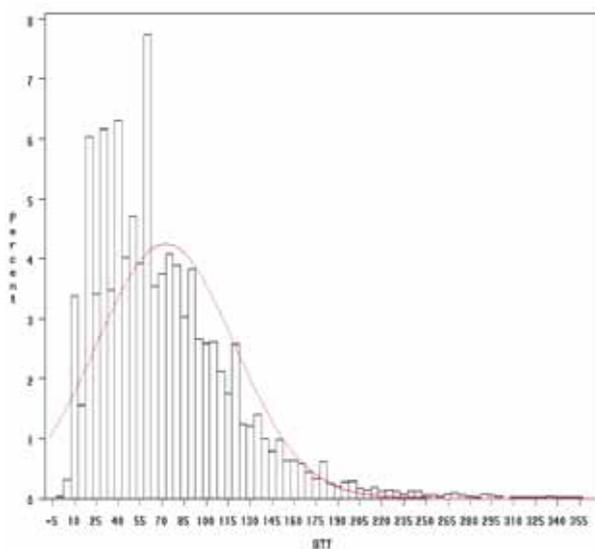
Paramètre	Symbole	Estimation
Moyenne	Mu	72.20575
Std Dev	Sigma	46.98612

### Tests d'ajustement pour distribution Normal

Test	Statistique	Seuil de significativité
<b>Kolmogorov-Smirnov</b>	D 0.110	Pr > D <0.010
<b>Cramer-von Mises</b>	W-Sq 33.557	Pr > W-Sq <0.005
<b>Anderson-Darling</b>	A-Sq 213.245	Pr > A-Sq <0.005
<b>Jarque-Bera</b>	Chi 7633.43	Pr > D <0.001

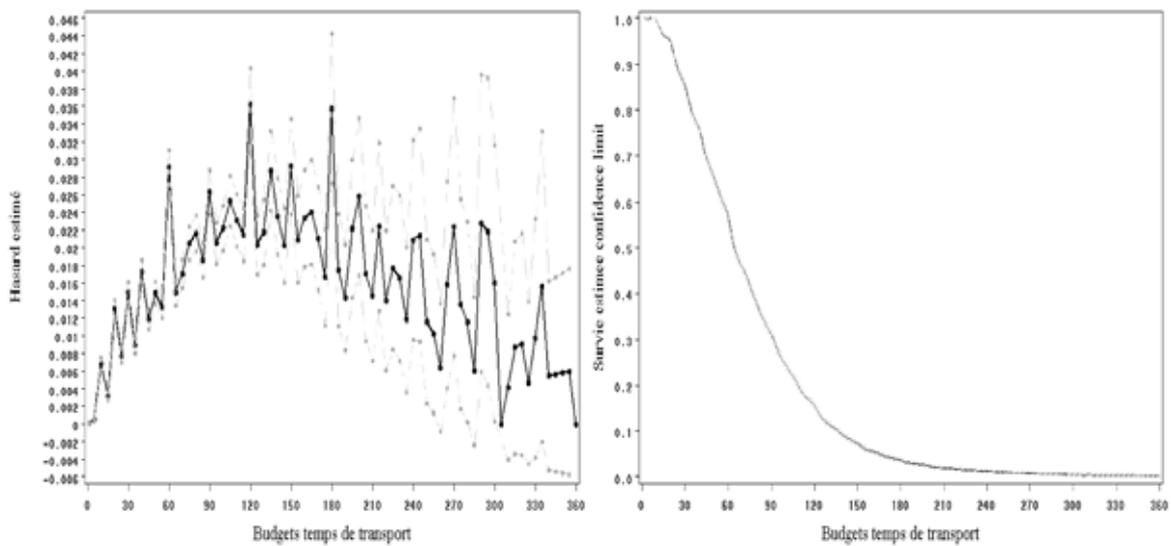
### Quantiles pour distribution Normal

Pourcentage	Quantile	
	Observé	Estimé
1.0	10.0	-37.100
5.0	15.0	-5.079
10.0	20.0	11.990
25.0	40.0	40.514
50.0	60.0	72.205
75.0	95.0	103.89
90.0	134.0	132.420
95.0	160.0	149.491
99.0	230.0	181.511

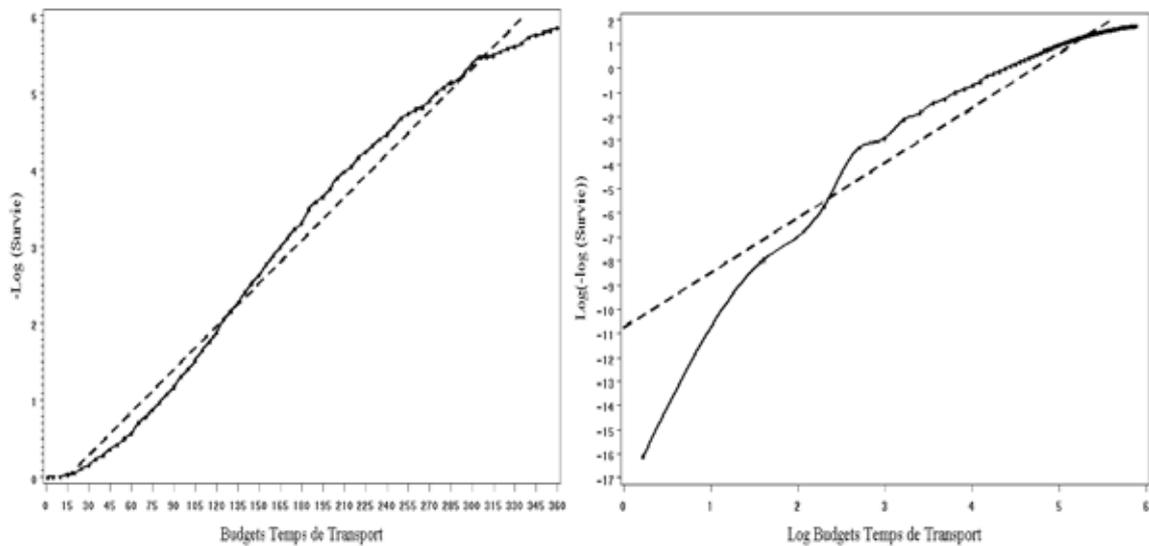


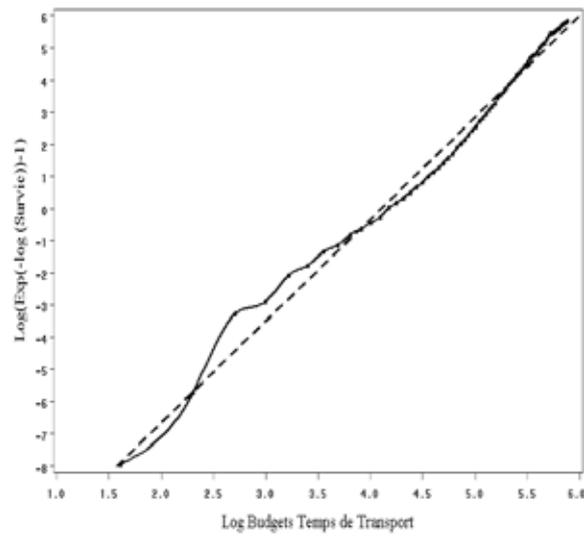
# Annexe XV - Hasard et survie – estimation non-paramétrique

## Hasard et survie estimés et leurs intervalles de confiance à 95%

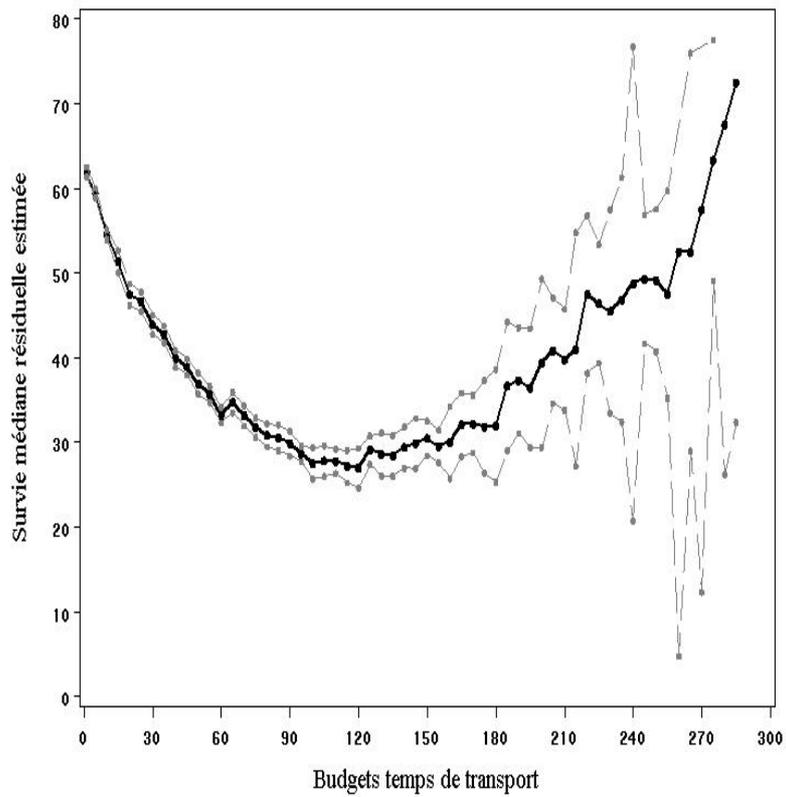


## Courbes des transformations du hasard intégré estimé (-Log(S) ; Log(-Log(S)) ; Log(Exp(-Log(S)))





**Survie médiane résiduelle estimée et intervalle de confiance à 95%**



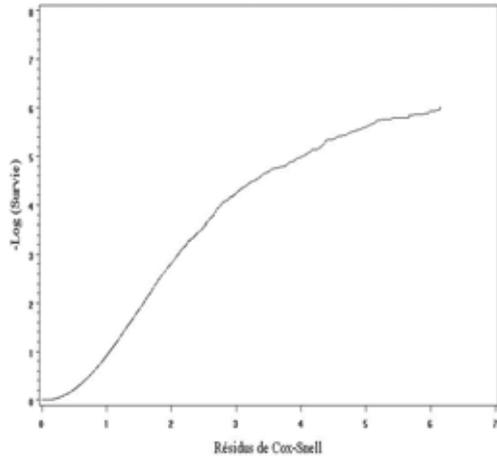
## Annexe XVI - Tests d'équivalence des survies

Variable	DF	Log Rank	Wilcoxon	-2Log(LR)
Nombre d'enfants	5	77.4202***	79.3792***	34.5027***
Nombre d'enfants de moins de 5 ans	2	7.3647**	4.1068	3.2709
Nombre d'enfants de plus de 6 ans	5	70.3744***	75.0428***	32.4598***
Présence d'enfants	1	31.4538***	24.0131***	14.5056***
Présence d'enfants de moins de 5 ans	1	6.1197**	4.1114**	2.6541
Nombre d'enfants (0 ; 1 ; 2+)	2	73.3784***	77.5265***	32.9934***
Possession du permis de conduire	1	124.7913***	219.0930***	77.2567***
Genre	1	29.2563***	19.4696***	16.3713***
Zone de domicile	7	33.8662***	37.6677***	16.5135**
Jour de déplacements	4	38.8173***	57.7420***	16.3304**
Statut professionnel en 9 classes	8	403.9847***	551.0887***	210.7751***
Statut Professionnel en 4 classes	3	361.4854***	492.1540***	186.6386***
Actif	1	126.3211***	231.8659***	84.0165***
Nombre d'automobile à disposition	4	24.3370***	28.8723***	12.7446**
Ménage motorisé	1	2.6504	0.0574	0.6564
Taille du ménage	8	25.1668***	28.0407***	13.2965
Revenu du ménage en 9 classes	10	16.4583*	41.6779***	8.6518
Revenu du ménage en 3 classes	2	10.4007***	30.0147***	5.1489*
Mode principal	6	1353.1074***	1995.4582***	579.7462***
Classe d'âge en 3 classes	2	243.7411***	346.7997***	131.0654***

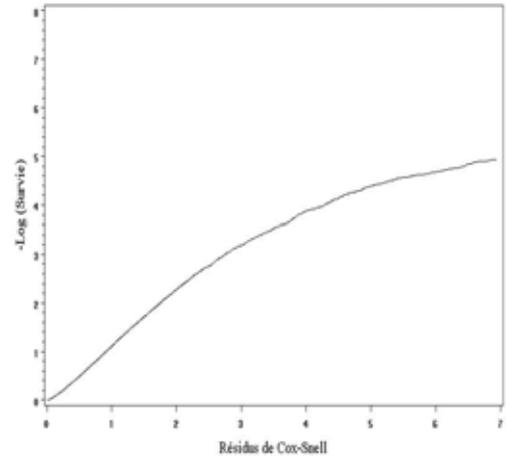
seuil de significativité : \* 0,1 ; \*\* 0,05 ; \*\*\* 0,01

## Annexe XVII - Résidus de Cox-Snell pour l'ensemble S3

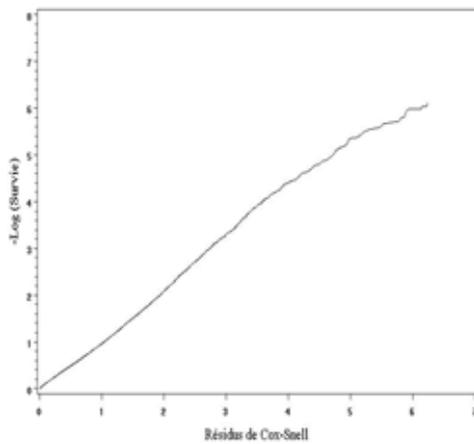
Exponentielle



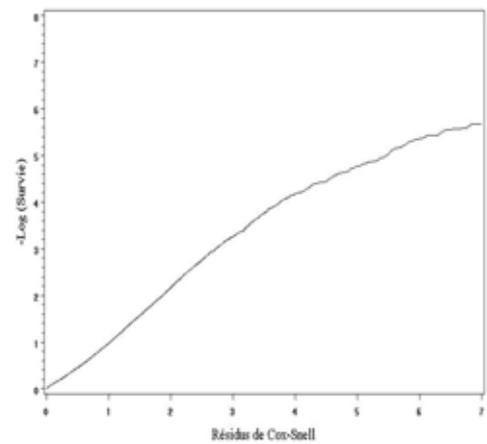
Weibull



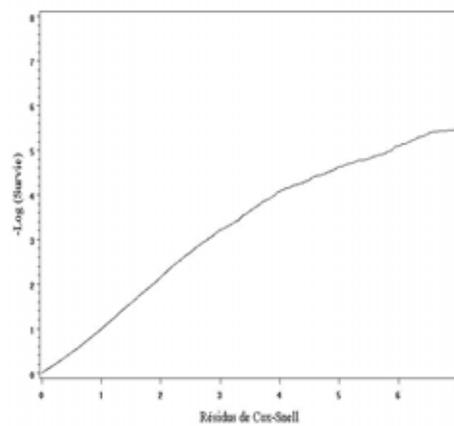
Log-logistique



Log-normale

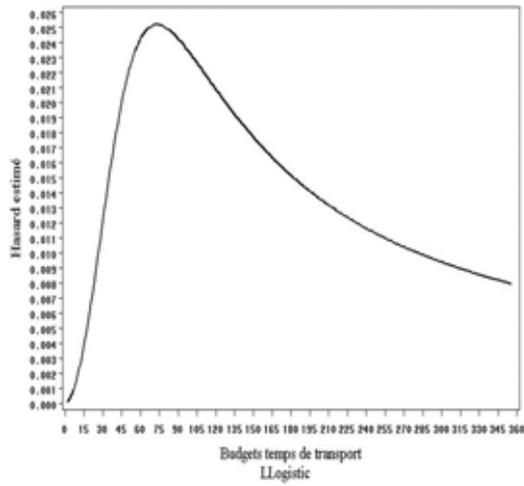


Gamma généralisée

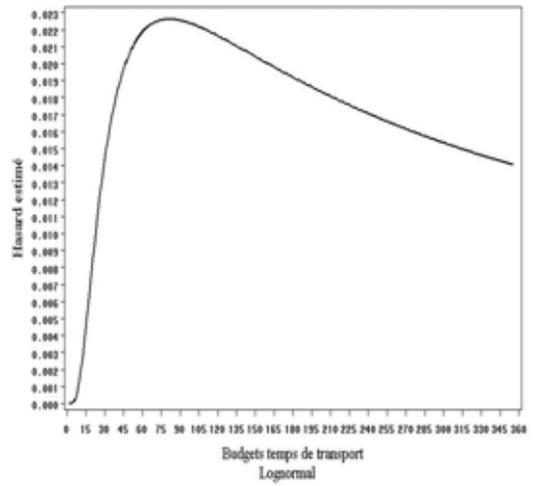


## Annexe XVIII - Hasards paramétriques estimés

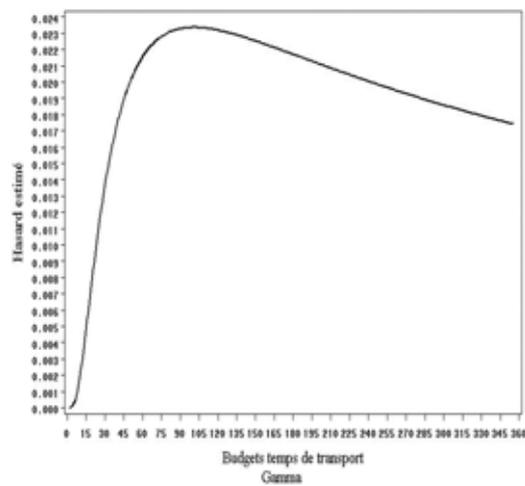
Log logistique



Log-normale



Gamma généralisée



## Annexe XIX - Estimation du modèle S3 avec une distribution Log-Normale

Analysis of Parameter Estimates

Parameter	DDL	Estimation	Erreur standard	95% Confidence Limits		Khi-2	Pr > ChiSq
Constante	1	3.8394	0.0421	3.7568	3.9220	8298.67	<.0001
homme	1	0.0572	0.0124	0.0330	0.0815	21.34	<.0001
Age	1	0.0012	0.0005	0.0002	0.0023	5.70	0.0170
Classe d'âge [0 ;20]	1	0.0856	0.0268	0.0332	0.1381	10.24	0.0014
Actif	1	0.2150	0.0220	0.1718	0.2581	95.22	<.0001
Nb d'enfants de plus de 5 ans	1	-0.0187	0.0096	-0.0376	0.0002	3.77	0.0520
Nb d'enfants de moins de 5 ans	1	-0.0625	0.0149	-0.0917	-0.0334	17.67	<.0001
Taille du ménage	1	0.0392	0.0071	0.0252	0.0531	30.26	<.0001
Ménage à haut revenu	1	0.0237	0.0171	-0.0098	0.0571	1.92	0.1655
Ménage à faible revenu	1	0.0278	0.0143	-0.0002	0.0557	3.78	0.0517
Centre-ville	1	0.0524	0.0221	0.0090	0.0958	5.61	0.0178
3eme couronne ouest	1	0.0405	0.0192	0.0029	0.0780	4.46	0.0348
3eme couronne est	1	-0.0559	0.0197	-0.0945	-0.0173	8.06	0.0045
Lundi	1	-0.0395	0.0148	-0.0685	-0.0105	7.12	0.0076
Vendredi	1	0.0410	0.0172	0.0074	0.0747	5.70	0.0169
Deux-roues motorisées	1	-0.6549	0.0255	-0.7050	-0.6049	658.88	<.0001
Marche	1	-0.6777	0.0779	-0.8304	-0.5251	75.73	<.0001
Modes publics	1	0.0959	0.0262	0.0445	0.1473	13.37	0.0003
Vélo	1	-0.2987	0.0237	-0.3452	-0.2522	158.50	<.0001
Modes privés (auto)	1	-0.3728	0.0250	-0.4218	-0.3238	222.45	<.0001
Temps de travail	1	0.0002	0.0000	0.0001	0.0003	18.94	<.0001
Temps de loisir	1	0.0015	0.0001	0.0013	0.0016	497.01	<.0001
Temps d'achat	1	0.0022	0.0001	0.0020	0.0025	267.43	<.0001
Accompagnement(0/1)	1	0.3019	0.0185	0.2656	0.3381	266.84	<.0001
Scale	1	0.6320	0.0042	0.6237	0.6403		
Log Likelihood		-10839.36					

## Annexe XX - Estimation du modèle S3 avec une distribution Gamma-généralisée

Analysis of Parameter Estimates

Parameter	DDL	Estimation	Erreur standard	95% Confidence Limits		Khi-2	Pr > ChiSq
Constante	1	3.8903	0.0424	3.8071	3.9735	8402.02	<.0001
Homme	1	0.0625	0.0123	0.0383	0.0866	25.67	<.0001
Age	1	0.0014	0.0005	0.0003	0.0024	6.89	0.0087
Classe d'âge [0 ;20]	1	0.0712	0.0266	0.0191	0.1234	7.16	0.0075
Actif	1	0.2245	0.0221	0.1812	0.2678	103.09	<.0001
Nb d'enfants de plus de 5 ans	1	-0.0208	0.0095	-0.0395	-0.0021	4.75	0.0293
Nb d'enfants de moins de 5ans	1	-0.0607	0.0148	-0.0897	-0.0318	16.89	<.0001
Taille du ménage	1	0.0394	0.0071	0.0255	0.0532	30.97	<.0001
Ménage à haut revenu	1	0.0253	0.0170	-0.0079	0.0585	2.23	0.1354
Ménage à faible revenu	1	0.0324	0.0142	0.0045	0.0602	5.20	0.0226
1ere couronne est	1	0.0526	0.0220	0.0095	0.0956	5.73	0.0167
3eme couronne ouest	1	0.0451	0.0190	0.0077	0.0824	5.60	0.0180
3eme couronne est	1	-0.0510	0.0196	-0.0893	-0.0126	6.79	0.0092
Lundi	1	-0.0405	0.0147	-0.0693	-0.0117	7.60	0.0058
Vendredi	1	0.0357	0.0171	0.0023	0.0692	4.38	0.0363
Deux-roues motorisées	1	-0.6321	0.0256	-0.6822	-0.5820	611.38	<.0001
Marche	1	-0.6595	0.0774	-0.8112	-0.5078	72.60	<.0001
Modes publics	1	0.0903	0.0260	0.0393	0.1414	12.02	0.0005
Vélo	1	-0.2866	0.0236	-0.3330	-0.2403	146.99	<.0001
Modes privés (auto)	1	-0.3539	0.0249	-0.4028	-0.3050	201.34	<.0001
Temps de travail	1	0.0001	0.0000	0.0000	0.0002	7.16	0.0075
Temps de loisir	1	0.0014	0.0001	0.0013	0.0015	432.92	<.0001
Temps d'achat	1	0.0021	0.0001	0.0018	0.0024	222.10	<.0001
Accompagnement(0/1)	1	0.2842	0.0185	0.2478	0.3205	235.29	<.0001
Scale	1	0.6275	0.0042	0.6193	0.6359		
Shape	1	0.1534	0.0210	0.1123	0.1945		
Log Likelihood		-10812.68					