

Université de Montréal

**Tests pour la dépendance entre les sections dans
un modèle de Poisson**

par

Arnaud Roussel

Département de mathématiques et de statistique

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)

en statistique

28 mai 2015

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

Tests pour la dépendance entre les sections dans un modèle de Poisson

présenté par

Arnaud Roussel

a été évalué par un jury composé des personnes suivantes :

Christian Léger

(président-rapporteur)

Jean-François Angers

(directeur de recherche)

Maciej Augustyniak

(membre du jury)

Mémoire accepté le

28 Mai 2015

SOMMAIRE

Pour des données de panel, les mesures répétées dans le temps peuvent remettre en cause l'hypothèse d'indépendance entre les individus. Des tests ont été développés pour pouvoir vérifier s'il reste de la dépendance entre les résidus d'un modèle. Les trois tests que nous présentons dans ce mémoire sont ceux de Pesaran (2004), Friedman (1937) et Frees (1995). Ces trois tests se basent sur les résidus (et leurs corrélations) et ont été construits pour des modèles linéaires. Nous voulons étudier dans ce mémoire les performances de ces trois tests dans le cadre d'un modèle linéaire généralisé de Poisson. Dans ce but, on compare tout d'abord leurs performances (niveaux et puissances) pour deux modèles linéaires, l'un ayant un terme autorégressif et l'autre non. Par la suite, nous nous intéressons à leurs performances pour un modèle linéaire généralisé de Poisson en s'inspirant de Hsiao, Pesaran et Pick (2007) qui adaptent le test de Pesaran (2004) pour un modèle linéaire généralisé. Toutes nos comparaisons de performances se feront à l'aide de simulations dans lesquelles nous ferons varier un certain nombre de paramètres (nombre d'observations, force de la dépendance, etc.). Nous verrons que lorsque les corrélations sont toutes du même signe, le test de Pesaran donne en général de meilleurs résultats, à la fois dans les cas linéaires et pour le modèle linéaire généralisé. Le test de Frees présentera de bonnes propriétés dans le cas où le signe des corrélations entre les résidus alterne.

Mots-clés : données de panel, résidus, régression, test, dépendance, modèle linéaire, modèle linéaire généralisé, Poisson

ABSTRACT

For panel data, repeated measures over time can challenge the hypothesis of dependence between subjects. Tests were developed in order to assess if some dependence remains among residuals. The three tests we present in this master thesis are from Pesaran (2004), Friedman (1937) and Frees (1995). These three tests, constructed specifically for linear models, are based on the residuals generated from models (and their correlations). We wish to study in this master thesis the performances of these three tests in the case of generalized linear Poisson models. For that goal, we compare them between each other (level, power, etc.) using two linear models, one with an autoregressive term and the other without. Next, inspired by Hsiao, Pesaran and Pick (2007) who adapt the test from Pesaran (2004), we will study their performances in a generalized Poisson model. All of our comparisons are done with simulations by modifying some variables (number of observations, strength of the dependence). We will observe that when the correlation is always of the same sign, Pesaran's test is the best in most cases, for the linear models and the generalized linear model. Frees' test will show good performances when the sign of the correlations alternates.

Keywords : panel data, residuals, regression, test, dependence, linear model, generalized linear model, Poisson

TABLE DES MATIÈRES

Sommaire.....	v
Abstract.....	vii
Liste des figures	xi
Liste des tableaux	xiii
Liste des abréviations et notations	1
Introduction.....	3
Chapitre 1. Statistiques de test pour l'indépendance des individus dans des données de panel.....	7
1.1. Test de Pesaran (2004).....	9
1.1.1. Présupposés et distribution asymptotique	11
1.2. Test de Friedman (1937)	12
1.3. Test de Frees (1995)	15
Chapitre 2. Simulations pour des modèles linéaires	21
2.1. Modèles de simulation	22
2.1.1. Modèle de Pesaran.....	22
2.1.2. Modèle simplifié	27
2.2. Résultats des simulations et comparaisons des tests	27

2.2.1. Modèle de Pesaran	27
2.2.2. Modèle simplifié	32
2.3. Résultats théoriques concernant les modèles	37
Chapitre 3. Simulations pour un modèle linéaire généralisé	49
3.1. Modèle de simulation	51
3.2. Résultats des simulations et comparaisons des tests	56
3.3. Résultats théoriques concernant le modèle.....	61
Conclusion.....	67
Bibliographie	69
Annexe A. Test des différents types de résidus pour modèle linéaire généralisé	A-i
Annexe B. Résumé des modèles utilisés pour les simulations.....	B-i
Annexe C. Test de Frees pour $T=5$.....	C-i

LISTE DES FIGURES

1.1	Comparaison entre loi normale et Q pour $T = 4$	18
1.2	Comparaison entre loi normale et Q pour $T = 5$	18
1.3	Comparaison entre loi normale et Q pour $T = 7$	18
1.4	Comparaison entre loi normale et Q pour $T = 10$	19

LISTE DES TABLEAUX

2.1	Niveaux, modèle linéaire de Pesaran $\gamma_i = 0$, test du <i>CD</i> de Pesaran . .	28
2.2	Niveaux, modèle linéaire de Pesaran $\gamma_i = 0$, test du <i>FR</i> de Friedman .	28
2.3	Niveaux, modèle linéaire de Pesaran $\gamma_i = 0$, test du <i>FRE</i> de Frees . . .	28
2.4	Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,1;0,3)$, test du <i>CD</i> de Pesaran	29
2.5	Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,1;0,3)$, test du <i>FR</i> de Friedman	29
2.6	Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,1;0,3)$, test du <i>FRE</i> de Frees	29
2.7	Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,2;0,4)$, test du <i>CD</i> de Pesaran	30
2.8	Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,2;0,4)$, test du <i>FR</i> de Friedman	30
2.9	Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,2;0,4)$, test du <i>FRE</i> de Frees	30
2.10	Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,3;0,5)$, test du <i>CD</i> de Pesaran	31
2.11	Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,3;0,5)$, test du <i>FR</i> de Friedman	31

2.12	Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du <i>FRE</i> de Frees	31
2.13	Niveaux, modèle linéaire simplifié $\gamma_i = 0$, test du <i>CD</i> de Pesaran	32
2.14	Niveaux, modèle linéaire simplifié $\gamma_i = 0$, test du <i>FR</i> de Friedman	32
2.15	Niveaux, modèle linéaire simplifié $\gamma_i = 0$, test du <i>FRE</i> de Frees	32
2.16	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,1;0,3)$, test du <i>CD</i> de Pesaran	33
2.17	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,1;0,3)$, test du <i>FR</i> de Friedman	33
2.18	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,1;0,3)$, test du <i>FRE</i> de Frees	33
2.19	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du <i>CD</i> de Pesaran	34
2.20	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du <i>FR</i> de Friedman	34
2.21	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du <i>FRE</i> de Frees	34
2.22	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du <i>CD</i> de Pesaran	35
2.23	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du <i>FR</i> de Friedman	35
2.24	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du <i>FRE</i> de Frees	35
2.25	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } N(0;1)$, test du <i>CD</i> de Pesaran	36

2.26	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } N(0;1)$, test du <i>FR</i> de Friedman	36
2.27	Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } N(0;1)$, test du <i>FRE</i> de Frees	36
3.1	Niveaux, modèle linéaire généralisé $\gamma_i = 0$, test du <i>CD</i> de Pesaran....	56
3.2	Niveaux, modèle linéaire généralisé $\gamma_i = 0$, test du <i>FR</i> de Friedman ..	56
3.3	Niveaux, modèle linéaire généralisé $\gamma_i = 0$, test du <i>FRE</i> de Frees.....	56
3.4	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,1;0,3)$, test du <i>CD</i> de Pesaran	57
3.5	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,1;0,3)$, test du <i>FR</i> de Friedman	57
3.6	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,1;0,3)$, test du <i>FRE</i> de Frees	57
3.7	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du <i>CD</i> de Pesaran	58
3.8	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du <i>FR</i> de Friedman	58
3.9	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du <i>FRE</i> de Frees	58
3.10	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du <i>CD</i> de Pesaran	59
3.11	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du <i>FR</i> de Friedman	59

3.12	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim Siid U(0,3;0,5)$, test du <i>FRE</i> de Frees	59
3.13	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim Siid N(0;1)$, test du <i>CD</i> de Pesaran	60
3.14	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim Siid N(0;1)$, test du <i>FR</i> de Friedman	60
3.15	Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim Siid N(0;1)$, test du <i>FRE</i> de Frees	60
A.1	Puissance, modèle linéaire généralisé, $\gamma_i \sim U(0,1;0,3)$. Résidus de type Réponse test du <i>CD</i>	A-ii
A.2	Puissance, modèle linéaire généralisé, $\gamma_i \sim U(0,1;0,3)$. Résidus de type Pearson test du <i>CD</i>	A-ii
A.3	Puissance, modèle linéaire généralisé, $\gamma_i \sim U(0,1;0,3)$. Résidus de type Déviance test du <i>CD</i>	A-ii
C.1	Niveaux, modèle linéaire de Pesaran $\gamma_i = 0$, comparaisons pour le test de Frees lorsque $T=5$	C-i
C.2	Niveaux, modèle linéaire simplifié $\gamma_i = 0$, comparaisons pour le test de Frees lorsque $T=5$	C-i
C.3	Puissance, modèle linéaire simplifié $\gamma_i \sim U(0,3;0,5)$, comparaisons pour le test de Frees lorsque $T=5$	C-ii

LISTE DES ABRÉVIATIONS ET NOTATIONS

CD	« Cross Dependence »
GLM	« Generalized Linear Model »
$N(\mu;\sigma)$	Distribution normale de moyenne μ et écart-type σ
$U(a;b)$	Distribution uniforme de paramètre a et b
iid	Indépendantes et identiquement distribuées
Siid	Simulées indépendantes et identiquement distribuées
$E(X)$	Espérance de X
$V(X)$	Variance de X
$Cov(X,Y)$	Covariance de X et Y

À moins d'avis contraire, la notation $\sim iid$ signifie que les variables aléatoires sont indépendantes et identiquement distribuées selon les indices inscrits. Par exemple, $\epsilon_{it} \sim iidN(0;1)$ signifie que les variables ϵ_{it} suivent pour tout indice i et t des lois normales indépendantes et identiquement distribuées de moyenne 0 et écart-type 1. La notation $\sim Siid$ sera utilisée pour indiquer que le paramètre est généré de façon aléatoire pour la génération de données de simulation et n'est pas un paramètre aléatoire du modèle ou une hypothèse faite pour les tests.

INTRODUCTION

L'analyse de données longitudinales est un mariage entre la régression et les séries chronologiques. La régression s'intéresse en général à une mesure par individu avec l'hypothèse que ceux-ci sont indépendants. Les séries chronologiques sont quant à elles très présentes en économétrie et représentent une mesure répétée dans le temps, comme par exemple le PIB d'un pays chaque année. Les données longitudinales sont l'étude de données sur plusieurs temps et pour plusieurs individus. On parle aussi dans la littérature francophone de données de panel ou de données croisées. Dans ce mémoire, ces trois termes sont interchangeables. L'hypothèse d'indépendance entre les individus propre à la régression peut être dans ce cas remise en question à cause des mesures répétées dans le temps pour chaque individu. Une dépendance entre les individus peut être due à des chocs communs entre les sujets, à l'omission de certaines variables qui entrent alors dans le terme d'erreur, à une dépendance spatiale, etc. Certains articles étudiant ce type de données proposent alors une modélisation de cette dépendance entre les différents individus constituant les données. Voir entre autres Anselin et Bera (1998), Sarafidis, Yamagata, et Robertson (2009), et Pesaran (2004) qui étudient ce sujet.

Pesaran (2004) et Frees (1995) s'intéressent à la question et proposent deux tests permettant de tester l'hypothèse d'indépendance entre les individus (aussi appelés sections) pour des données de panel linéaires. Dans Frees (1995), il est également proposé d'utiliser une statistique provenant de Friedman (1937) afin

de tester cette même hypothèse. C'est cette utilisation de la statistique proposée dans Friedman (1937) que nous présentons dans ce mémoire comme étant le test de Friedman. Ces trois tests s'intéressent aux résidus générés par l'estimation du modèle et à leurs corrélations deux à deux. Le test de Pesaran utilise la corrélation de Pearson alors que les deux autres la corrélation non paramétrique de Spearman. Ces tests se basent sur une valeur moyenne de ces corrélations qui devrait suivre une certaine distribution sous l'hypothèse de corrélation nulle. En effet, il est commun dans la littérature de tester l'hypothèse nulle d'indépendance en testant si la corrélation est nulle (voir par exemple Frees 2004 section 2.17). Nous proposons donc dans ce mémoire d'étudier les performances de ces trois tests pour des modèles linéaires. Nous utilisons deux modèles linéaires pour nos comparaisons, l'un présentant un terme autorégressif comme dans Pesaran (2004) et l'autre sans. Hsiao, Pesaran, et Pick (2007) s'intéressent à l'efficacité du test de Pesaran (2004) dans le cas où le modèle est un modèle linéaire généralisé. Ils réalisent des simulations sur des modèles *probit* et *logit*. Les résultats qu'ils obtiennent montrent que le test de Pesaran (2004) présente de bonnes propriétés aussi dans ce contexte. Nous proposons donc dans ce mémoire de comparer également les trois tests avec un modèle linéaire généralisé de Poisson. Il est nécessaire dans ce contexte de faire un choix sur la forme des résidus. Nous choisirons d'utiliser les résidus de Pearson comme il est courant pour ce type de régression (voir Fox et Sanford 2011). Frees (2004) note que les tests de Pesaran et de Friedman devraient être moins puissants lorsque les corrélations entre les différents résidus alternent de signe. Nous comparons donc les trois tests dans les deux cas, lorsque le signe est toujours le même (positif dans ce mémoire) ou lorsqu'il alterne. Ceci sera fait à la fois pour les modèles linéaires considérés et pour le modèle linéaire généralisé.

Dans le premier chapitre de ce mémoire, nous présentons les trois tests qui seront utilisés et comparés par la suite. Nous présentons brièvement dans quel

contexte les statistiques de test utilisées ont été développées et quelles sont leurs distributions. Nous verrons que Pesaran (2004) s'intéresse principalement aux avantages que propose sa statistique comparativement à celle proposée dans Breusch et Pagan (1980). Il montre ceci par des simulations d'un modèle que nous reprendrons par la suite. Frees (2004) réutilise une statistique proposée par Friedman (1937) pour pouvoir tester l'hypothèse d'indépendance des individus. C'est cette utilisation que nous nommons test de Friedman et nous verrons les différentes caractéristiques de la statistique dans la section qui lui est consacrée. Enfin, nous finirons ce chapitre par la statistique de Frees (1995) pour laquelle nous indiquons la convergence de la statistique ainsi que les possibles approximations pouvant être faites pour réaliser le test.

Dans le deuxième chapitre, nous introduisons les différents modèles linéaires qui sont simulés et auxquels nous appliquons les trois tests vus au chapitre un. Le premier modèle linéaire est celui de Pesaran (2004). Ce modèle introduit une forme du terme d'erreur permettant de générer de la dépendance entre les individus du modèle. Cette forme est ensuite conservée pour nos autres modèles de simulation. Nous changeons simplement la distribution d'une des variables pour pouvoir comparer les trois tests lorsque le signe de la corrélation alterne de signe comme suggéré par Frees (2004). Le modèle dans Pesaran (2004) utilise un terme d'autorégression. Nous étudions également les performances des trois tests dans un modèle, dit simplifié, sans ce terme d'autorégression. Les résultats de simulation des trois tests sont ensuite exposés pour leurs niveaux et leurs puissances selon différentes formes et forces pour la dépendance. Dans la suite du chapitre, nous développons également des formules théoriques sur les différentes covariances entre les individus pour les deux modèles. Ceci nous permet de montrer que la forme de l'erreur suggérée par Pesaran (2004) génère bien de la dépendance entre

les individus uniquement.

Enfin, dans le troisième et dernier chapitre, nous réalisons un travail similaire au chapitre deux mais en nous inspirant de Hsiao *et al.* (2007). Nous appliquons les trois tests à des données provenant d'un modèle linéaire généralisé de Poisson. Nous exposons donc dans un premier temps le modèle utilisé ainsi que les étapes de simulation. Par la suite, les résultats de niveaux et de puissances sont présentés. Dans la dernière section de ce chapitre, nous exposons une approximation des covariances de notre modèle linéaire généralisé en utilisant la méthode delta.

L'ensemble des simulations de ce mémoire sont réalisées avec 1000 itérations et sont faites pour un nombre d'observations allant de 5 à 300 et pour un nombre de temps de 5 à 300.

Chapitre 1

STATISTIQUES DE TEST POUR L'INDÉPENDANCE DES INDIVIDUS DANS DES DONNÉES DE PANEL

Dans ce premier chapitre, on introduit les modèles pour données de panel (de manière générale). On voit ensuite dans les sections suivantes les trois différents tests qui seront utilisés et comparés dans la suite du mémoire. Un modèle linéaire pour données de panel prend la forme suivante :

$$y_{it} = \alpha_i + \beta_i' x_{it} + u_{it} \text{ avec } i = 1, 2, \dots, N \text{ et } t = 1, 2, \dots, T, \quad (1.0.1)$$

où i est l'indice de l'individu, t est l'indice de temps. N est donc le nombre d'individus et T le nombre de mesures de temps. Le vecteur x_{it} représente les variables explicatives mesurées dans le temps pour l'unité i au temps t . Le modèle assumé donne donc la variable réponse y_{it} en fonction des variables explicatives avec les paramètres de régression α_i et β_i (vecteur des variables explicatives) selon une relation linéaire. Ces derniers peuvent dépendre des sujets tout comme ils peuvent être fixes selon le modèle choisi. Pour i fixé (c'est-à-dire pour un même individu), les erreurs u_{it} sont indépendantes et identiquement distribuées, de moyenne 0 et de variance σ_i^2 . On note également que les β_i et α_i sont fonctions de i mais non de

t . Comme le note Pesaran (2004), la condition selon laquelle les u_{it} sont indépendantes dans le temps est non restrictive. En effet, l'on pourrait modifier le modèle en prenant en compte cette dépendance en ajoutant suffisamment de termes autorégressifs du type $y_{i(t-k)}$. Cependant, une dépendance entre les individus peut exister (dépendance entre u_{it} et u_{jt} pour $i \neq j$) pour diverses raisons (variables communes non prises en compte, groupes de sujets ayant une dépendance, chocs communs, etc.). Voir par exemple Baltagi, Song, Jung, et Koh (2007) ou Baltagi (2005) sections 10.5 et 12.3. On peut également trouver une éventuelle solution pour pallier au problème en corrigeant les erreurs comme dans Driscoll et Kraay (1998).

Sous l'hypothèse où les individus sont indépendants, on voit que la covariance entre u_{it} et u_{jt} est nulle pour tout $i \neq j$. On peut donc poser notre hypothèse nulle pour la suite du mémoire :

$$H_0 : \text{Les erreurs } u_{it} \text{ sont indépendantes,} \quad (1.0.2)$$

ce qui implique que dans ce cas $\text{Cov}(u_{it}, u_{jt}) = 0$ pour tout i et j si $i \neq j$. Il est indiqué entre autres dans Pesaran (2004) que l'approche habituelle est de tester l'hypothèse de covariance nulle. Comme on le voit par la suite dans les trois sections suivantes, les trois tests utilisés dans ce mémoire proposent des statistiques pour tester cette hypothèse et se basent sur les résidus du modèle. Si on note $\hat{\alpha}_i$ et $\hat{\beta}_i$ les estimateurs des paramètres mentionnés plus haut, posons :

$$e_{it} = y_{it} - \hat{\alpha}_i - \hat{\beta}_i' x_{it}, \quad (1.0.3)$$

les résidus de la régression linéaire (estimateurs de u_{it}). Les statistiques de test utilisent une sommation de l'ensemble des différents estimateurs des corrélations entre les erreurs u_{it} et u_{jt} de deux individus différents (deux à deux). Pesaran (2004) utilise la corrélation usuelle de Pearson alors que les deux autres tests

utilisent la corrélation de Spearman. La distribution de ces statistiques est alors analysée sous l'hypothèse d'indépendance (impliquant une covariance nulle) afin de pouvoir réaliser le test. La modélisation utilisée dans le mémoire pour intégrer de la dépendance entre les individus est décrite au chapitre deux.

1.1. TEST DE PESARAN (2004)

Dans cette première section, on expose le test introduit dans Pesaran (2004). Le test utilise une statistique abrégée CD pour « Cross Dependence ». La formule de cette statistique dépend des corrélations entre les résidus de chaque ligne du modèle. Les résidus sont définis dans l'équation (1.0.2). On prend un estimateur de la corrélation entre ces résidus, $\hat{\rho}_{ij}$:

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^T e_{it}e_{jt}}{\sqrt{\sum_{t=1}^T e_{it}^2 \sum_{t=1}^T e_{jt}^2}}, \quad (1.1.1)$$

ou encore :

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \sum_{t=1}^T \xi_{it}\xi_{jt},$$

avec

$$\xi_{it} = \frac{e_{it}}{\sqrt{E_i' E_i}}$$

et $E_i = (e_{i1}, e_{i2}, \dots, e_{iT})'$.

La formule de la statistique CD est alors la suivante :

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij} \right). \quad (1.1.2)$$

Cette statistique fait donc intervenir la somme des différentes corrélations entre les résidus de chaque section. Le terme devant la sommation des corrélations

est simplement un terme de standardisation. Sous certaines conditions exposées à la section 1.1.1 et sous H_0 (équation 1.1.1), cette statistique suit asymptotiquement en N une loi normale standard ($N(0; 1)$). Par conséquent, l'hypothèse nulle d'indépendance est rejetée lorsque la valeur absolue du CD est trop grande (avec un niveau de confiance à 95% on rejette pour des valeurs absolues supérieures à 1,96).

Un des avantages soulevés par Pesaran (2004) est que cette statistique ne dépend pas du présupposé *a priori* d'une matrice de dépendance spatiale (matrice de poids). Dans certains cas, lorsque l'information est disponible, on peut supposer une dépendance dans le temps et/ou l'espace entre les observations (par exemple, si les individus représentent des sujets pris dans différents endroits, il est possible que ceux-ci soient spatialement corrélés). On utilise alors une matrice de poids $\mathbf{W} = (w_{it})$ apparaissant dans la modélisation des résidus : $u_{it} = \lambda \sum_{j=1}^N w_{ij} u_{jt} + \epsilon_{it}$. Cette formulation montre bien qu'on y suppose que u_{it} dépend des autres u_{jt} avec des poids w_{ij} . On peut aussi trouver cette définition sous forme matricielle $\mathbf{U} = \lambda \mathbf{W} \mathbf{U} + \epsilon$. Pour plus de détails sur les processus spatiaux et les définitions possibles de la matrice de poids, voir Cliff et Ord (1981). La forme que doit prendre cette dépendance est en général impliquée par le type de données utilisées dans un contexte. Un test sur la présence de cette dépendance reviendrait à tester $\lambda = 0$ (voir Anselin et Bera 1998). Dans Pesaran (2004), on cherche cependant une méthode qui *a priori* ne dépend pas de cette dépendance spatiale. Une statistique plus ancienne par Breusch et Pagan (1980) est également exposée :

$$CD_{lm} = \sqrt{\frac{1}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N (T \hat{\rho}_{ij}^2 - 1) \right). \quad (1.1.3)$$

Cependant, cette statistique a besoin de T suffisamment grand pour que $T \hat{\rho}_{ij}^2 \sim \chi_1^2$ (asymptotiquement) sous l'hypothèse nulle ; on peut retrouver ce résultat dans Breusch et Pagan (1980) et Pesaran (2004). Il y est indiqué que asymptotiquement

en T , $\sqrt{T}\hat{\rho}_{ij} \sim N(0;1)$ dont le carré est χ_1^2 . Si cette condition est respectée (T assez grand), alors la loi de CD_{lm} converge vers une $N(0,1)$ avec N . Pour T fixé $E(T\hat{\rho}_{ij}^2 - 1)$ n'est pas exactement égal à 0 (et plus T est petit, plus c'est inexact). Ainsi, si T est trop petit alors cette approximation est grossière et l'erreur générée se retrouve amplifiée si N est important en comparaison. C'est pourquoi Pesaran (2004) propose sa statistique du CD qui donne de meilleurs résultats dans les cas où N est beaucoup plus grand que T car sa statistique, fortement inspirée de celle du CD_{lm} , est exactement centrée en 0 (sous l'hypothèse nulle). Pesaran (2004) obtient de meilleurs résultats numériques pour le CD que le CD_{lm} ce qui soutient cet argument.

1.1.1. Présupposés et distribution asymptotique

Maintenant que le contexte et la statistique du CD ont été introduits, on rappelle ici les conditions nécessaires à la statistique du CD telles qu'exposées par Pesaran (2004). La statistique CD nécessite les présupposés suivants :

- (1) Pour tout i , les erreurs u_{it} sont indépendantes (par rapport au temps) et de variance σ_i^2 .
- (2) Sous l'hypothèse nulle : $u_{it} = \sigma_i \epsilon_{it}$ avec $\epsilon_{it} \sim iid$ de variance 1 et symétriquement centrés en 0 pour tout i et tout t .
- (3) Les régresseurs sont strictement exogènes, c'est-à-dire extérieurs au modèle (qui ne dépend pas des autres variables du modèle). On a également $E(u_{it}|X_i) = 0$ pour tout i et tout t .
- (4) $T > k + 1$, k étant le nombre de régresseurs, et les résidus de la régression des moindres carrés ne sont pas tous nuls.

Selon ces présupposés et sous l'hypothèse nulle (Théorème 1 de Pesaran 2004) :

$$E(\hat{\rho}_{ij}) = 0,$$

et

$$E(CD) = 0.$$

Sous H_0 et avec ces présupposés, on peut montrer (voir Pesaran 2004) que lorsque N tend vers l'infini pour T fixé, la statistique CD tend vers une $N(0,1)$. Il y est fait la conjecture qu'étant donné que l'espérance du CD est exactement 0, la statistique devrait avoir de bonnes propriétés même pour de petits N et T contrairement à la statistique CD_{lm} qui elle est centrée asymptotiquement en 0. En effet, comme on l'a vu plus haut, le CD_{lm} se base sur une distribution asymptotique ce qui risque de créer des problèmes lorsque N est beaucoup plus grand que T puisque le biais engendré par de petits T est multiplié proportionnellement au nombre d'observations N . Dans cette situation, le CD_{lm} est trop grand et on a tendance à rejeter l'hypothèse nulle plus souvent. Ce phénomène est visible d'après les simulations dans Pesaran (2004) où le CD_{lm} échoue parfois complètement dans certaines situations (par exemple avec indépendance, $N = 100$, $T = 5$ le niveau est à 1).

1.2. TEST DE FRIEDMAN (1937)

On expose ici le test proposé dans Friedman (1937) dans la forme présentée dans Frees (2004) et Hoyos et Sarafidis (2006). Friedman (1937) développe à l'origine sa statistique, notée FR ici, pour une méthode qu'il nomme méthode des rangs. Pour des données sous-divisées selon un facteur, le test permet de déterminer si ce facteur a une influence significative sur la variable sous-divisée. Par exemple, dans son article, il utilise des données sur les dépenses des ménages dans différentes catégories (éducation, transport, etc.). Les familles sont ensuite sous-divisées en fonction de leur revenu et on s'intéresse aux écart-types pour chaque combinaison de niveau de revenu et catégorie de dépenses. Il cherche ensuite à déterminer si ces écart-types varient significativement selon le niveau de revenu. Il calcule alors les rangs de ces écart-types pour chaque catégorie. S'il n'y a pas

d'effet significatif on s'attendrait donc à ce que pour chaque niveau de revenu les rangs des différentes catégories soient aléatoires. L'article consacre aussi une partie à comparer cette méthode à celle de l'analyse de variance et souligne qu'un des avantages est de n'avoir besoin que des rangs et ainsi des présupposés moins restrictifs sur la nature des variables (normalité non requise par exemple). Voir aussi Brown et Mood (1951). Cette méthode, et la statistique FR associée, est aussi utilisée dans le contexte où N individus rangent T objets et où l'on cherche à déterminer s'il y a une différence entre les objets (s'ils sont, ou non, rangés aléatoirement) comme par exemple dans Sen (1967).

Dans Frees (2004), le test est adapté au contexte que nous voyons ici où l'on cherche à déterminer s'il existe une dépendance entre les individus des données de panel à partir des résidus. La statistique FR utilise la corrélation non paramétrique de Spearman. Celle-ci est en réalité la corrélation usuelle de Pearson (comme utilisée au chapitre un) appliquée aux rangs des deux vecteurs. Si on note $(r_{i,1}, r_{i,2}, \dots, r_{i,T})$ et $(r_{j,1}, r_{j,2}, \dots, r_{j,T})$ les vecteurs des rangs des résidus respectifs $(e_{i,1}, e_{i,2}, \dots, e_{i,T})$ et $(e_{j,1}, e_{j,2}, \dots, e_{j,T})$, alors la corrélation de Spearman est calculée comme étant la corrélation échantillonnale de ces rangs :

$$\hat{\rho}_{S,ij} = \hat{\rho}_{S,ji} = \frac{\sum_{t=1}^T \left(r_{it} - \left[\frac{T+1}{2} \right] \right) \left(r_{jt} - \left[\frac{T+1}{2} \right] \right)}{\sqrt{\sum_{t=1}^T \left(r_{it} - \left[\frac{T+1}{2} \right] \right)^2 \sum_{t=1}^T \left(r_{jt} - \left[\frac{T+1}{2} \right] \right)^2}}. \quad (1.2.1)$$

La valeur $\frac{T+1}{2}$ est le rang moyen des deux vecteurs (la moyenne de $(1, \dots, T)$). Une formule simplifiée de cette corrélation peut aussi être utilisée :

$$\hat{\rho}_{S,ij} = 1 - \frac{6 \sum_{t=1}^T d_t^2}{T(T^2 - 1)} \quad (1.2.2)$$

où d_t est la différence de rang pour un même $t : r_{it} - r_{jt}$. Cette formule est exacte si on n'a pas de rangs égaux (lorsque les variables sous-jacentes sont égales). Comme dans ce mémoire nous travaillons avec des résidus continus nous n'avons pas ce problème (au chapitre 3 les données sont discrètes mais les résidus sont continus). Pour plus de détails sur cette corrélation, voir par exemple Hollander et Wolfe (1973) section 8.5. La statistique FR est alors :

$$FR = (T - 1)[(N - 1)R_M + 1] \quad (1.2.3)$$

où

$$R_M = \frac{2}{N(N - 1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{S,ij}, \quad (1.2.4)$$

avec $\hat{\rho}_{S,ij}$ la corrélation de Spearman comme vue plus haut. On trouve également une autre formulation possible, plus économe en temps de calcul, de cette même statistique dans Friedman (1937) :

$$FR = \frac{12}{NT(T + 1)} \sum_{t=1}^T \left(\sum_{i=1}^N r_{it} \right)^2 - 3N(T + 1). \quad (1.2.5)$$

Cette formulation est plus économe en coût computationnel car elle ne nécessite pas de calculer les $N(N - 1)/2$ corrélations. Comme cette corrélation est non-paramétrique, Frees (1995) note qu'utiliser une telle corrélation permet de se défaire des hypothèses habituelles sur la distribution des erreurs (normalité, valeurs aberrantes) si on est en présence d'un modèle ne satisfaisant pas ces exigences.

D'après Friedman (1937), sous les conditions H_0 d'indépendance et donc de covariance nulle, FR suit asymptotiquement en N une chi-deux à $T - 1$ degrés de liberté pour T fixé. Le test revient donc à rejeter l'hypothèse nulle lorsque FR est supérieure au quantile d'une chi-deux à $T - 1$ degrés de liberté (quantile calculé selon le niveau de confiance choisi).

1.3. TEST DE FREES (1995)

Dans Frees (2004), il y est souligné que les deux tests vus précédemment (Pesaran 2004 et Friedman 1937) peuvent présenter des problèmes lorsque le signe de la corrélation utilisée dans les deux cas alterne d'un individu à l'autre. Autrement dit, puisque les deux statistiques utilisent une sommation de corrélations sans en prendre le carré ni la valeur absolue, il peut y avoir une perte de puissance dans le cas où on a beaucoup de corrélations positives et beaucoup de corrélations négatives. Un exemple de cette situation est décrit au chapitre deux au moment de faire les simulations. Il est alors suggéré d'utiliser une autre statistique que nous notons FRE dans ce mémoire. Elle est exposée dans Frees (1995).

Cette statistique utilise le carré des corrélations de Spearman comme vu à la section précédente. Frees (1995) utilise comme présupposé (sous l'hypothèse nulle d'indépendance) qu'il faut simplement que les erreurs soient *iid* et symétriques. On note tout d'abord :

$$R_M^2 = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{S,ij}^2. \quad (1.3.1)$$

On peut voir ici encore une fois qu'intuitivement ceci est la moyenne des différentes corrélations entre les résidus des individus pris deux à deux. L'étude de cette quantité est faite dans Frees (1995). Il y est notamment montré que si les erreurs ont une densité continue

$$FRE = N \left(R_M^2 - \frac{1}{T-1} \right) \quad (1.3.2)$$

converge avec N vers une variable

$$Q = a(T) \left(\chi_{1,T-1}^2 - (T-1) \right) + b(T) \left(\chi_{2,T(T-3)/2}^2 - \frac{T(T-3)}{2} \right), \quad (1.3.3)$$

où $\chi_{1,T-1}^2$ et $\chi_{2,T(T-3)/2}^2$ sont deux variables aléatoires de chi-deux indépendantes de degrés de liberté $T-1$ et $\frac{T(T-3)}{2}$ respectivement, $a(T) = \frac{4(T+2)}{5(T-1)^2(T+1)}$, et $b(T) = \frac{2(5T+6)}{5T(T-1)(T+1)}$. Frees (1995) donne également les formules pour la variance et l'espérance de R_M^2 :

$$E(R_M^2) = \frac{1}{T-1},$$

$$V(R_M^2) = \binom{N}{2}^{-1} \frac{(T-2)(25T^2 - 7T - 54)}{18(T-1)^2(T^3 - T)}.$$

Réaliser le test demande donc de calculer les quantiles de la variable Q pour trouver les valeurs critiques du test. Ceci pouvant être compliqué, Frees (2004) suggère d'utiliser une approximation normale qui s'avère bonne pour T plus grand que 10. Il est facile de voir que $E(Q) = 0$ (on rappelle que l'espérance d'une chi-deux est son degré de liberté). En rappelant que la variance d'une chi-deux de degré k est $2k$ on a directement la variance de Q :

$$V(Q) = \frac{32}{25} \frac{(T+2)^2}{(T-1)^3(T+1)^2} + \frac{4}{25} \frac{(5T+6)^2(T-3)}{T(T-1)^2(T+1)^2} \quad (1.3.4)$$

et l'approximation consiste donc à prendre pour T assez grand $\frac{FRE}{\sqrt{V(Q)}} \sim N(0;1)$ et donc de rejeter l'hypothèse nulle lorsque $\frac{FRE}{\sqrt{V(Q)}}$ est plus grand en valeur absolue que 1,96 par exemple pour un test au niveau 5%. On note aussi que la formule de la variance de Q donnée dans Hoyos et Sarafidis (2006) comporte une erreur, le deuxième terme est divisé par 25 et non par 5.

Nous avons également vérifié jusqu'à quel point cette approximation de Q par une normale était valide. Pour cela, nous nous sommes tout d'abord intéressés à la densité de la variable Q . Afin de simplifier la lecture, nous notons $a = a(T)$, $b = b(T)$, $X = \chi_{1,T-1}^2$ et $Y = \chi_{2,T(T-3)/2}^2$ les différentes variables présentes dans Q . Nous pouvons développer Q :

$$Q = c + aX + bY$$

en prenant $c = -a \times (T - 1) - b \times \frac{T(T-3)}{2}$. En posant $R = Y$, nous pouvons donc déterminer la densité conjointe de Q et R à partir de celle de X et Y . Il est assez évident de voir que les transformations inverses sont

$$X = \frac{Q - c - bR}{a} \text{ et } Y = R.$$

La matrice jacobienne de ces fonctions inverses est alors

$$J = \begin{pmatrix} \frac{1}{a} & \frac{-b}{a} \\ 0 & 1 \end{pmatrix}$$

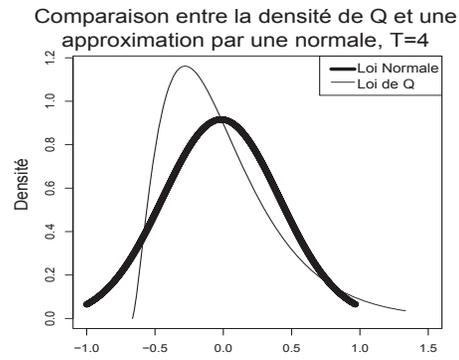
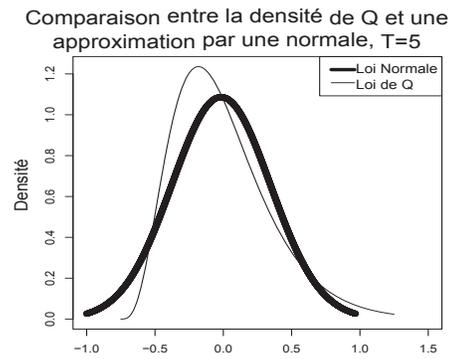
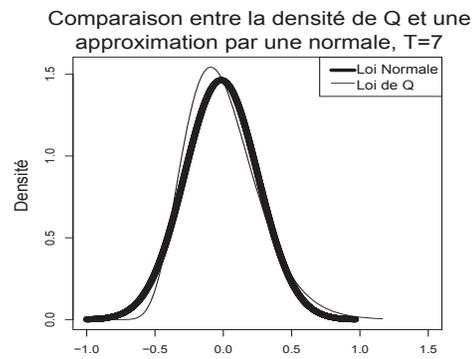
et donc de déterminant $\frac{1}{a}$. La fonction de densité conjointe de Q et R sera donc

$$f_{Q,R}(q, r) = \frac{1}{a} f_X \left(\frac{q - c - br}{a} \right) f_Y(r),$$

où f_X et f_Y sont les densités des deux chi-deux X et Y . La variable X devant être positive, la variable $Y = R$ est donc bornée entre 0 et $\frac{Q-c}{b}$. Donc la densité de Q en q est le résultat de l'intégrale suivante :

$$f_Q(q) = \int_0^{\frac{q-c}{b}} \frac{1}{a} f_X \left(\frac{q - c - br}{a} \right) f_Y(r) dr. \quad (1.3.5)$$

Comme ces fonctions sont celles de deux densités de chi-deux de degrés différents, nous ne voyons pas de solution analytique simple et avons donc calculé cette intégrale numériquement. Voici quatre graphiques pour $T = 4, T = 5, T = 7$ et $T = 10$ comparant la densité de la loi pour Q et celle de la loi normale de moyenne 0 et de variance $V(Q)$. En augmentant T ($T \geq 10$) les deux densités deviennent très similaires.

FIGURE 1.1. Comparaison entre loi normale et Q pour $T = 4$ FIGURE 1.2. Comparaison entre loi normale et Q pour $T = 5$ FIGURE 1.3. Comparaison entre loi normale et Q pour $T = 7$

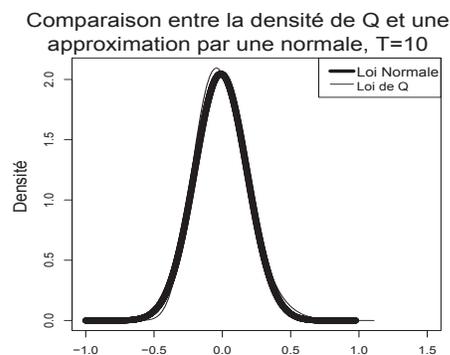


FIGURE 1.4. Comparaison entre loi normale et Q pour $T = 10$

Maintenant que les trois tests qui sont utilisés dans le mémoire ont été introduits nous nous occupons de les comparer pour des modèles linéaires dans un premier temps puis pour un modèle de régression de Poisson ensuite. Les différents modèles utilisés sont introduits dans les chapitres respectifs ainsi que dans l'annexe B.

Chapitre 2

SIMULATIONS POUR DES MODÈLES LINÉAIRES

Dans ce chapitre, nous introduisons les deux modèles de simulations qui sont utilisés. Le premier provient de Pesaran (2004) et comporte un terme d'autorégression. Les erreurs sont modélisées de manière à rendre compte de la dépendance à tester. Nous effectuons également des simulations avec un modèle dit simplifié en réutilisant la même modélisation des erreurs. Frees (2004) souligne que les tests de Pesaran et Friedman peuvent présenter des problèmes de puissance si les corrélations alternent de signe. Nous comparons donc également les tests dans cette situation.

Une fois les modèles introduits, nous présentons dans la même section les étapes détaillées des simulations, puis dans la section suivante les résultats de niveaux et de puissance pour les trois tests. Les trois tests sont réalisés simultanément sur une même itération du modèle et nous réalisons 1000 itérations pour chaque combinaison des variables. Pour 1000 itérations, la marge d'erreur de niveau 95% due à la simulation est de 1,35% pour le calcul du niveau d'un test à 5%.

2.1. MODÈLES DE SIMULATION

Dans cette section on décrit les modèles de Pesaran (2004) que nous notons, y_{it}^A ainsi que le modèle simplifié noté y_{it}^L . Nous introduisons également la formulation de la dépendance entre les individus qui est utilisée dans les deux cas.

2.1.1. Modèle de Pesaran

La section suivante est consacrée aux simulations faites dans Pesaran (2004). Le modèle choisi est

$$y_{it}^A = \mu_i(1 - \beta_i) + \beta_i y_{i(t-1)}^A + u_{it}, \quad (2.1.1)$$

où la variable β_i est générée selon une $U(0;1)$ et μ_i selon une $N(0;2)$. On remarque ici que ce modèle ne suit pas le présupposé numéro 3 décrit dans la section 1.1.1 puisque $y_{i(t-1)}$ est endogène. Pesaran (2004) se justifie tout de même d'utiliser ce modèle en notant que son test continue de donner des résultats satisfaisants. C'est pour cette raison que nous avons décidé d'étudier également le comportement des tests sans ce terme d'autorégression grâce au modèle simplifié décrit dans la sous-section suivante. Le terme d'erreur u_{it} dépend de la force de la dépendance souhaitée. Les erreurs sont simulées à l'aide du modèle suivant :

$$u_{it} = \gamma_i f_t + \epsilon_{it}, i = 1, 2, \dots, N; t = 1, 2, \dots, T. \quad (2.1.2)$$

Le facteur $f_t \sim N(0;1)$ est l'élément permettant de modéliser la dépendance entre les individus. Comme on le verra dans les résultats théoriques la covariance entre les erreurs et les individus est due à ce facteur. Les erreurs ϵ_{it} sont supposés suivre des normales centrées en 0 et sont générées avec une variance de 1 pour les simulations. Dans le cas de l'indépendance entre les individus nous avons $\gamma_i = 0$. Dans le cas alternatif γ_i permet de moduler la force de la dépendance pour les simulations. Comme Pesaran (2004), nous utilisons le scénario de simulation suivant :

- (1) Faible dépendance : $\gamma_i \sim \text{Siid } U(0,1;0,3)$,
- (2) Moyenne dépendance : $\gamma_i \sim \text{Siid } U(0,2;0,4)$,
- (3) Forte dépendance : $\gamma_i \sim \text{Siid } U(0,3;0,5)$.

Ce processus de génération des erreurs est réutilisé dans l'ensemble du mémoire. La notation $\sim \text{Siid}$, qui sera souvent utilisée, permet de souligner que le paramètre est généré de façon aléatoire uniquement pour la génération de données des simulations et ne représente pas une supposition du modèle nécessaire à l'inférence et à l'utilisation des tests. Nous réalisons donc des simulations pour chaque force de dépendance. Le niveau du test consiste à rejeter l'hypothèse d'indépendance lorsque nous sommes dans le cas de l'hypothèse nulle, c'est à dire $\gamma_i = 0$. Cependant il est aisé de voir que pour un même t , γ_i étant strictement positive, les erreurs générées pour la simulation sont toujours corrélées positivement. Nous montrons dans la section 2.3 que la covariance entre u_{it} et u_{jt} est toujours positive ou nulle. C'est ce que Frees (1995) souligne en notant que si au contraire γ_i est généré à partir d'une loi centrée en 0, par exemple $\gamma_i \sim \text{Siid } N(0, \sigma_\gamma)$, alors en reprenant les mêmes démarches que celles de la section 2.3 nous allons alors générer une covariance nulle entre u_{it} et u_{jt} . Mais on aura pourtant une dépendance ($\text{Cov}(u_{it}, u_{jt} | \gamma_i, \gamma_j) \neq 0$) à cause du facteur f_t . Pour cette raison nous faisons également une comparaison entre les trois tests en utilisant $\gamma_i \sim \text{Siid } N(0;1)$ dans les simulations pour le modèle simplifié pour obtenir des corrélations qui alternent de signe. Puisque cela ne concerne que des simulations pour la puissance des tests, il n'est bien entendu pas nécessaire de refaire des comparaisons de niveaux puisque le modèle est inchangé sous l'hypothèse nulle.

Pour résumer, voici les différentes étapes en détail de nos simulations dans le cadre du modèle de Pesaran (2004) :

Étape 1 : Générer y_{it}^A

Les variables f_t et ϵ_{it} sont simulés à partir de $N(0;1)$ et la variable μ_i est simulée avec une $N(0;2)$. La variable β_i est simulée avec une uniforme(0;1). La variable simulant la dépendance γ_i suit une certaine distribution uniforme de paramètres variant selon le degré de dépendance souhaité (voir précédemment). Pour vérifier les performances du test de Frees, nous prenons également γ_i simulé à partir d'une normale standard. Ensuite en fixant $y_{i0} = 0$ on construit :

$$y_{it}^A = \mu_i(1 - \beta_i) + \beta_i y_{i(t-1)}^A + u_{it}$$

avec

$$u_{it} = \gamma_i f_t + \epsilon_{it}.$$

Étape 2 : Faire les régressions

Pour chaque i , on réalise une régression linéaire de y_{it}^A selon $y_{i(t-1)}^A, t = 1, \dots, T$ avec une ordonnée à l'origine. Cette estimation ignore la présence du facteur f_t puisque le but est de voir si le test détecte une dépendance entre les résidus dans ce cas. Les résidus de ces régressions notés e_{it} sont ensuite calculés ainsi que la corrélation usuelle de Pearson entre e_{it} et e_{jt} pour $t = 1, 2, \dots, T$. Nous calculons également la corrélation de Spearman entre ces mêmes résidus (en calculant tout d'abord le rang des résidus puis la corrélation de ces rangs). On note que les résidus utilisés ici sont les résidus ordinaires de la régression par les moindres carrés ($e_{it} = y_{it}^A - \hat{y}_{it}^A$ où \hat{y}_{it}^A sont les valeurs prédites de la régression).

Étape 3 : Calcul des différentes statistiques de test et décision

Si $\hat{\rho}_{ij}$ est la corrélation usuelle de Pearson entre les résidus e_{it} et e_{jt} alors on calcule :

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij} \right).$$

Sous l'hypothèse nulle, le CD converge vers une normale standard. Donc si la valeur absolue du CD calculée est supérieure à la valeur critique de 1,96 alors on détecte de la dépendance pour cette itération (pour un taux d'erreur à 5%).

Si $\hat{\rho}_{S,ij}$ est la corrélation de Spearman alors nous calculons les deux choses suivantes :

$$R_M = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{S,ij}$$

et

$$R_M^2 = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{S,ij}^2.$$

À partir de ces quantités on peut donc calculer les statistiques de Friedman et Frees qui sont respectivement :

$$FR = (T-1)[(N-1)R_M + 1]$$

et

$$FRE = N \left(R_M^2 - \frac{1}{T-1} \right).$$

La quantité FR converge en N vers une chi-deux à $T-1$ degrés de liberté sous l'hypothèse nulle. Nous comparons donc la valeur FR au quantile 95% d'une chi-deux à $T-1$ degrés de liberté et rejetons l'hypothèse nulle si FR est plus grand.

La quantité FRE converge vers la variable Q décrite dans la section 1.3. Cependant nous utiliserons l'approximation suggérée par Frees (2004) en considérant

que $\frac{FRE}{\sqrt{V(Q)}}$ converge vers une normale standard. Donc en calculant

$$V(Q) = \frac{32}{25} \frac{(T+2)^2}{(T-1)^3(T+1)^2} + \frac{4}{25} \frac{(5T+6)^2(T-3)}{T(T-1)^2(T+1)^2}$$

nous aurons $\frac{FRE}{\sqrt{V(Q)}}$ et rejetterons si la valeur absolue est supérieure à 1,96. On note qu'il est possible que le test présente des problèmes lorsque $T = 5$ car l'approximation de Q en loi normale y est plus grossière.

On rappelle que la puissance du test est sa capacité (en probabilité) à détecter de la dépendance lorsque celle-ci est présente. On souhaiterait que celle-ci soit le plus proche de 100% (ou 1) possible. Lorsque $\gamma_i = 0$, on teste en fait si le test détecte de la dépendance par erreur (on parle d'erreur de type 1), dans ce cas on parle du niveau du test et cela devrait correspondre au niveau d'erreur choisi (ici 5%). L'objectif de ces simulations est d'obtenir ces valeurs de puissances et de niveaux sous certaines conditions pour voir si le test génère de bons résultats. Pour 1000 itérations, les valeurs de niveaux calculées ont une marge d'erreur de 1,35%.

2.1.2. Modèle simplifié

Nous désirons également simuler les trois tests sans le terme d'autorégression $y_{i(t-1)}^A$ présent dans le test de Pesaran. Dans cette sous-section, nous introduisons donc un modèle simplifié pour lequel nous allons également utiliser le *CD* de Pesaran, *FR* de Friedman, et *FRE* de Frees. Le modèle simplifié prend l'expression suivante :

$$y_{it}^L = \alpha_i + \beta_i x_{it} + u_{it}. \quad (2.1.3)$$

Les erreurs u_{it} sont simulées de la même façon que précédemment (équation 2.1.2). Pour les autres variables, nous prenons $\alpha_i \sim \text{Siid } U(0;1)$, $\beta_i \sim \text{Siid } U(0;1)$ et $x_{it} \sim \text{Siid } N(2,1)$. Ce modèle plus simple ne présente plus de terme d'autorégression ce qui nous permet de voir l'impact de ce terme sur les tests. Les différentes étapes de simulations sont identiques à celles de la sous-section précédente en ce qui concerne les étapes 2 et 3. Seule l'étape 1 est modifiée en utilisant le modèle et les distributions des différentes variables qui viennent d'être décrites ici.

2.2. RÉSULTATS DES SIMULATIONS ET COMPARAISONS DES TESTS

Dans cette section, nous présentons les différents résultats des simulations pour les trois tests. Les différentes étapes de simulation sont décrites dans la section précédente. Nous commençons par présenter les résultats pour le modèle de Pesaran. Chaque simulation est réalisée avec 1000 répétitions pour chaque combinaison de N , T et γ_i .

2.2.1. Modèle de Pesaran

Pour le modèle de Pesaran, voici les différents résultats pour les niveaux.

TABLE 2.1. Niveaux, modèle linéaire de Pesaran $\gamma_i = 0$, test du CD de Pesaran

N\T	5	10	25	50	100	200	300
5	0,066	0,054	0,046	0,041	0,041	0,047	0,033
10	0,069	0,048	0,052	0,052	0,061	0,053	0,056
25	0,071	0,060	0,044	0,044	0,045	0,053	0,050
50	0,075	0,061	0,040	0,045	0,053	0,058	0,043
100	0,076	0,063	0,054	0,068	0,049	0,043	0,059
200	0,060	0,048	0,047	0,057	0,035	0,050	0,048
300	0,088	0,057	0,046	0,053	0,050	0,045	0,049

TABLE 2.2. Niveaux, modèle linéaire de Pesaran $\gamma_i = 0$, test du FR de Friedman

N\T	5	10	25	50	100	200	300
5	0,041	0,035	0,032	0,037	0,028	0,038	0,023
10	0,059	0,031	0,041	0,040	0,043	0,054	0,039
25	0,056	0,053	0,042	0,044	0,039	0,050	0,046
50	0,057	0,051	0,045	0,052	0,059	0,053	0,041
100	0,059	0,057	0,057	0,062	0,045	0,044	0,056
200	0,049	0,045	0,052	0,050	0,048	0,049	0,041
300	0,065	0,068	0,046	0,045	0,061	0,045	0,040

TABLE 2.3. Niveaux, modèle linéaire de Pesaran $\gamma_i = 0$, test du FRE de Frees

N\T	5	10	25	50	100	200	300
5	0,083	0,076	0,071	0,065	0,048	0,068	0,047
10	0,089	0,078	0,053	0,066	0,057	0,070	0,065
25	0,165	0,105	0,053	0,048	0,050	0,060	0,052
50	0,368	0,175	0,054	0,058	0,058	0,057	0,058
100	0,749	0,438	0,093	0,046	0,037	0,055	0,052
200	0,978	0,874	0,146	0,078	0,063	0,062	0,049
300	0,999	0,992	0,251	0,080	0,047	0,057	0,050

La marge d'erreur pour une proportion de 5% et 1000 simulations est d'environ 1,35% (avec un niveau de confiance à 95%). On s'attend donc à des niveaux entre 3,65% et 6,35%. On remarque que le test du CD est légèrement au-delà de cet intervalle pour les cas où $T = 5$. Par contre, le test de Frees présente de grosses anomalies lorsque T est petit par rapport à N (principalement lorsque T est deux fois plus petit que N). Il est intéressant de remarquer que ce résultat est similaire à ceux qu'obtient Pesaran (2004) avec le CD_{lm} . Il le justifie par

une erreur d'approximation entre la loi asymptotique de la statistique et sa loi exacte dans un échantillon de taille finie. Cette approximation est par la suite amplifiée par le nombre d'observations. En utilisant le vrai quantile de la loi pour Q proposé dans Frees (2004) pour $T = 5, 0,683$, nous n'avons obtenu qu'une légère amélioration comme décrit en annexe C. Enfin, le test de Friedman présente quelques anomalies lorsque $N = 5$ (niveaux trop bas). Nous présentons maintenant les tableaux pour les trois tests pour une même force de dépendance sur une même page. Le nombre de simulations est le même.

TABLE 2.4. Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,1;0,3)$, test du CD de Pesaran

N\T	5	10	25	50	100	200	300
5	0,097	0,082	0,112	0,191	0,273	0,402	0,539
10	0,133	0,142	0,288	0,410	0,633	0,855	0,951
25	0,265	0,401	0,717	0,917	0,994	1,000	1,000
50	0,424	0,683	0,972	0,999	1,000	1,000	1,000
100	0,673	0,915	1,000	1,000	1,000	1,000	1,000
200	0,840	0,989	1,000	1,000	1,000	1,000	1,000
300	0,921	0,999	1,000	1,000	1,000	1,000	1,000

TABLE 2.5. Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,1;0,3)$, test du FR de Friedman

N\T	5	10	25	50	100	200	300
5	0,071	0,066	0,092	0,165	0,237	0,398	0,530
10	0,102	0,119	0,276	0,413	0,652	0,861	0,955
25	0,211	0,362	0,705	0,912	0,996	1,000	1,000
50	0,346	0,638	0,961	0,999	1,000	1,000	1,000
100	0,594	0,897	1,000	1,000	1,000	1,000	1,000
200	0,782	0,983	1,000	1,000	1,000	1,000	1,000
300	0,901	0,997	1,000	1,000	1,000	1,000	1,000

TABLE 2.6. Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim Siid U(0,1;0,3)$, test du FRE de Frees

N\T	5	10	25	50	100	200	300
5	0,091	0,061	0,087	0,100	0,130	0,215	0,270
10	0,106	0,072	0,082	0,101	0,163	0,336	0,517
25	0,177	0,111	0,108	0,199	0,440	0,799	0,971
50	0,386	0,218	0,200	0,434	0,816	0,999	1,000
100	0,747	0,531	0,484	0,796	0,988	1,000	1,000
200	0,969	0,927	0,834	0,984	1,000	1,000	1,000
300	0,998	0,997	0,959	0,993	1,000	1,000	1,000

TABLE 2.7. Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du *CD* de Pesaran

N\T	5	10	25	50	100	200	300
5	0,145	0,160	0,268	0,447	0,667	0,893	0,961
10	0,216	0,332	0,637	0,879	0,990	0,999	1,000
25	0,456	0,755	0,974	1,000	1,000	1,000	1,000
50	0,686	0,936	0,999	1,000	1,000	1,000	1,000
100	0,863	0,992	1,000	1,000	1,000	1,000	1,000
200	0,952	0,998	1,000	1,000	1,000	1,000	1,000
300	0,986	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 2.8. Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du *FR* de Friedman

N\T	5	10	25	50	100	200	300
5	0,080	0,109	0,218	0,403	0,626	0,881	0,964
10	0,163	0,267	0,614	0,868	0,985	1,000	1,000
25	0,377	0,702	0,972	1,000	1,000	1,000	1,000
50	0,636	0,912	0,999	1,000	1,000	1,000	1,000
100	0,826	0,985	1,000	1,000	1,000	1,000	1,000
200	0,939	0,998	1,000	1,000	1,000	1,000	1,000
300	0,974	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 2.9. Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du *FRE* de Frees

N\T	5	10	25	50	100	200	300
5	0,097	0,073	0,124	0,205	0,348	0,642	0,780
10	0,107	0,097	0,175	0,343	0,648	0,942	0,991
25	0,214	0,176	0,375	0,777	0,983	1,000	1,000
50	0,411	0,395	0,689	0,966	1,000	1,000	1,000
100	0,753	0,761	0,933	1,000	1,000	1,000	1,000
200	0,974	0,977	0,998	1,000	1,000	1,000	1,000
300	0,998	1,000	0,999	1,000	1,000	1,000	1,000

TABLE 2.10. Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du *CD* de Pesaran

N\T	5	10	25	50	100	200	300
5	0,183	0,276	0,491	0,773	0,944	0,997	1,000
10	0,334	0,573	0,895	0,988	1,000	1,000	1,000
25	0,612	0,901	0,999	1,000	1,000	1,000	1,000
50	0,831	0,982	1,000	1,000	1,000	1,000	1,000
100	0,939	1,000	1,000	1,000	1,000	1,000	1,000
200	0,979	0,999	1,000	1,000	1,000	1,000	1,000
300	0,992	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 2.11. Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du *FR* de Friedman

N\T	5	10	25	50	100	200	300
5	0,105	0,213	0,418	0,749	0,927	0,995	1,000
10	0,258	0,505	0,882	0,990	1,000	1,000	1,000
25	0,534	0,884	1,000	1,000	1,000	1,000	1,000
50	0,781	0,980	1,000	1,000	1,000	1,000	1,000
100	0,918	1,000	1,000	1,000	1,000	1,000	1,000
200	0,979	0,999	1,000	1,000	1,000	1,000	1,000
300	0,989	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 2.12. Puissance, modèle linéaire de Pesaran avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du *FRE* de Frees

N\T	5	10	25	50	100	200	300
5	0,090	0,122	0,215	0,428	0,726	0,971	0,997
10	0,121	0,156	0,408	0,745	0,972	1,000	1,000
25	0,246	0,366	0,805	0,988	1,000	1,000	1,000
50	0,491	0,629	0,954	1,000	1,000	1,000	1,000
100	0,776	0,904	0,999	1,000	1,000	1,000	1,000
200	0,976	0,995	1,000	1,000	1,000	1,000	1,000
300	0,998	1,000	1,000	1,000	1,000	1,000	1,000

Cette simulation de la puissance du test *CD* dans le modèle linéaire suit les résultats de Pesaran (2004). Les deux autres tests se comportent de manière similaire. En effet, la puissance est plus forte en fonction de γ_i et croît avec N et T et de manière plus rapide avec N que T . Le *CD* de Pesaran présente les meilleurs résultats avec le test non paramétrique de Friedman. Il faut noter que le test de Frees dans ce modèle de Pesaran n'a pas les bons niveaux lorsque T est beaucoup plus petit que N et donc, comme le test rejette trop souvent l'hypothèse nulle,

les données relatives à la puissance doivent être nuancées dans ce cas. De plus, même lorsque T est suffisamment grand, le test de Frees a une puissance moindre par rapport aux deux autres.

2.2.2. Modèle simplifié

Pour le modèle simplifié, on a les résultats suivants pour les niveaux.

TABLE 2.13. Niveaux, modèle linéaire simplifié $\gamma_i = 0$, test du CD de Pesaran

N\T	5	10	25	50	100	200	300
5	0,057	0,056	0,060	0,040	0,046	0,040	0,050
10	0,065	0,049	0,056	0,059	0,050	0,058	0,055
25	0,072	0,059	0,051	0,051	0,054	0,044	0,050
50	0,060	0,047	0,057	0,058	0,034	0,056	0,052
100	0,076	0,056	0,062	0,045	0,054	0,056	0,044
200	0,082	0,043	0,055	0,050	0,046	0,058	0,046
300	0,073	0,049	0,051	0,050	0,048	0,054	0,060

TABLE 2.14. Niveaux, modèle linéaire simplifié $\gamma_i = 0$, test du FR de Friedman

N\T	5	10	25	50	100	200	300
5	0,043	0,028	0,047	0,027	0,028	0,026	0,024
10	0,046	0,036	0,039	0,042	0,041	0,048	0,048
25	0,049	0,051	0,049	0,048	0,041	0,047	0,046
50	0,045	0,047	0,055	0,054	0,050	0,047	0,047
100	0,056	0,057	0,060	0,049	0,049	0,042	0,046
200	0,056	0,040	0,049	0,046	0,043	0,059	0,044
300	0,054	0,050	0,042	0,052	0,048	0,043	0,052

TABLE 2.15. Niveaux, modèle linéaire simplifié $\gamma_i = 0$, test du FRE de Frees

N\T	5	10	25	50	100	200	300
5	0,085	0,069	0,080	0,057	0,065	0,062	0,052
10	0,045	0,059	0,066	0,066	0,059	0,045	0,055
25	0,049	0,054	0,075	0,063	0,044	0,042	0,068
50	0,047	0,049	0,051	0,047	0,050	0,049	0,051
100	0,041	0,047	0,050	0,060	0,059	0,051	0,056
200	0,049	0,053	0,052	0,056	0,055	0,044	0,049
300	0,049	0,038	0,053	0,046	0,051	0,055	0,051

Dans ce modèle, le niveau est bien maintenu autour de 5% dans la plupart des cas et ce même pour le test de Frees avec ce modèle. Ceci laisserait supposer que

les problèmes obtenus avec le modèle de Pesaran pour le test de Frees pourraient être dus au terme d'autorégression. La marge d'erreur avec 1000 simulations étant la même (1,35%), on remarquera cependant des problèmes similaires pour le CD pour de petites valeurs de T ($T = 5$ principalement) et lorsque $N = 5$ pour le test de Frees. Les tableaux suivants de puissance des tests sont présentés trois par trois pour une même force de dépendance.

TABLE 2.16. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim Siid U(0,1;0,3)$, test du CD de Pesaran

N\T	5	10	25	50	100	200	300
5	0,093	0,102	0,135	0,167	0,219	0,411	0,546
10	0,132	0,157	0,264	0,408	0,644	0,870	0,963
25	0,245	0,400	0,717	0,907	0,992	1,000	1,000
50	0,387	0,678	0,959	0,998	1,000	1,000	1,000
100	0,629	0,910	1,000	1,000	1,000	1,000	1,000
200	0,822	0,986	1,000	1,000	1,000	1,000	1,000
300	0,907	0,997	1,000	1,000	1,000	1,000	1,000

TABLE 2.17. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim Siid U(0,1;0,3)$, test du FR de Friedman

N\T	5	10	25	50	100	200	300
5	0,044	0,062	0,106	0,157	0,209	0,394	0,530
10	0,098	0,118	0,252	0,401	0,643	0,873	0,965
25	0,169	0,376	0,705	0,908	0,995	1,000	1,000
50	0,328	0,641	0,957	0,999	1,000	1,000	1,000
100	0,561	0,894	0,999	1,000	1,000	1,000	1,000
200	0,769	0,982	1,000	1,000	1,000	1,000	1,000
300	0,873	0,997	1,000	1,000	1,000	1,000	1,000

TABLE 2.18. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim Siid U(0,1;0,3)$, test du FRE de Frees

N\T	5	10	25	50	100	200	300
5	0,068	0,065	0,073	0,086	0,098	0,202	0,285
10	0,047	0,064	0,061	0,101	0,197	0,349	0,517
25	0,051	0,054	0,100	0,190	0,437	0,821	0,969
50	0,051	0,076	0,155	0,385	0,809	0,993	1,000
100	0,051	0,109	0,374	0,772	0,995	1,000	1,000
200	0,071	0,166	0,642	0,973	1,000	1,000	1,000
300	0,082	0,268	0,818	0,997	1,000	1,000	1,000

TABLE 2.19. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du *CD* de Pesaran

N\T	5	10	25	50	100	200	300
5	0,116	0,148	0,268	0,431	0,660	0,886	0,973
10	0,207	0,333	0,620	0,845	0,988	1,000	1,000
25	0,440	0,745	0,974	0,999	1,000	1,000	1,000
50	0,663	0,939	1,000	1,000	1,000	1,000	1,000
100	0,835	0,991	1,000	1,000	1,000	1,000	1,000
200	0,947	0,999	1,000	1,000	1,000	1,000	1,000
300	0,972	0,999	1,000	1,000	1,000	1,000	1,000

TABLE 2.20. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du *FR* de Friedman

N\T	5	10	25	50	100	200	300
5	0,068	0,103	0,228	0,402	0,646	0,878	0,963
10	0,152	0,288	0,580	0,833	0,981	1,000	1,000
25	0,358	0,692	0,968	0,999	1,000	1,000	1,000
50	0,611	0,909	1,000	1,000	1,000	1,000	1,000
100	0,799	0,989	1,000	1,000	1,000	1,000	1,000
200	0,936	0,999	1,000	1,000	1,000	1,000	1,000
300	0,965	0,999	1,000	1,000	1,000	1,000	1,000

TABLE 2.21. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du *FRE* de Frees

N\T	5	10	25	50	100	200	300
5	0,063	0,058	0,109	0,167	0,336	0,650	0,817
10	0,049	0,067	0,169	0,303	0,649	0,951	0,991
25	0,063	0,102	0,374	0,759	0,981	1,000	1,000
50	0,082	0,201	0,651	0,961	1,000	1,000	1,000
100	0,097	0,366	0,893	0,999	1,000	1,000	1,000
200	0,173	0,574	0,983	1,000	1,000	1,000	1,000
300	0,213	0,708	0,995	1,000	1,000	1,000	1,000

TABLE 2.22. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du *CD* de Pesaran

N \ T	5	10	25	50	100	200	300
5	0,190	0,292	0,519	0,760	0,951	0,997	1,000
10	0,305	0,551	0,903	0,991	1,000	1,000	1,000
25	0,622	0,912	0,999	1,000	1,000	1,000	1,000
50	0,820	0,988	1,000	1,000	1,000	1,000	1,000
100	0,934	0,998	1,000	1,000	1,000	1,000	1,000
200	0,983	1,000	1,000	1,000	1,000	1,000	1,000
300	0,992	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 2.23. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du *FR* de Friedman

N \ T	5	10	25	50	100	200	300
5	0,104	0,209	0,475	0,728	0,944	0,997	1,000
10	0,214	0,494	0,885	0,992	1,000	1,000	1,000
25	0,559	0,887	0,998	1,000	1,000	1,000	1,000
50	0,762	0,981	1,000	1,000	1,000	1,000	1,000
100	0,913	0,999	1,000	1,000	1,000	1,000	1,000
200	0,975	1,000	1,000	1,000	1,000	1,000	1,000
300	0,986	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 2.24. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du *FRE* de Frees

N \ T	5	10	25	50	100	200	300
5	0,076	0,112	0,233	0,433	0,780	0,967	0,996
10	0,056	0,116	0,400	0,770	0,974	1,000	1,000
25	0,084	0,308	0,792	0,990	1,000	1,000	1,000
50	0,127	0,466	0,953	1,000	1,000	1,000	1,000
100	0,215	0,688	1,000	1,000	1,000	1,000	1,000
200	0,337	0,886	1,000	1,000	1,000	1,000	1,000
300	0,447	0,939	1,000	1,000	1,000	1,000	1,000

Les résultats sur la puissance sont très similaires à ceux de la sous-section précédente. Le *CD* présente les meilleurs résultats suivi par le test de Friedman. On note que le test de Frees a ici les bons niveaux et donc la comparaison des puissances est possible. Le test de Frees montre cependant une faible puissance lorsque $T = 5$. Comme nous avons ici utilisé l'approximation par la loi normale ceci peut expliquer les performances de la statistique *FRE* ici lorsque T est faible. Cependant en refaisant nos simulations pour $T = 5$ avec le vrai quantile pour la

loi Q nous n'obtenons qu'une très légère amélioration (voir annexe C). Comme expliqué dans la section 2.1.1, le test de Frees peut présenter un avantage lorsque γ_i est centré en 0. Nous réalisons donc une série de trois tableaux de puissance pour les tests en prenant $\gamma_i \sim N(0;1)$.

TABLE 2.25. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } N(0;1)$, test du CD de Pesaran

N\T	5	10	25	50	100	200	300
5	0,071	0,101	0,212	0,355	0,522	0,693	0,745
10	0,072	0,087	0,174	0,315	0,462	0,711	0,740
25	0,064	0,082	0,167	0,291	0,505	0,680	0,745
50	0,071	0,081	0,155	0,262	0,479	0,675	0,759
100	0,078	0,097	0,143	0,268	0,463	0,663	0,753
200	0,063	0,086	0,141	0,284	0,498	0,673	0,747
300	0,077	0,073	0,150	0,259	0,512	0,696	0,760

TABLE 2.26. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } N(0;1)$, test du FR de Friedman

N\T	5	10	25	50	100	200	300
5	0,045	0,073	0,127	0,173	0,220	0,245	0,279
10	0,050	0,077	0,109	0,162	0,171	0,242	0,240
25	0,051	0,077	0,124	0,158	0,178	0,222	0,257
50	0,053	0,072	0,115	0,155	0,184	0,213	0,247
100	0,067	0,078	0,107	0,149	0,186	0,224	0,218
200	0,053	0,086	0,114	0,150	0,206	0,225	0,240
300	0,074	0,065	0,119	0,139	0,218	0,234	0,235

TABLE 2.27. Puissance, modèle linéaire simplifié avec dépendance $\gamma_i \sim \text{Siid } N(0;1)$, test du FRE de Frees

N\T	5	10	25	50	100	200	300
5	0,111	0,348	0,725	0,897	0,964	0,988	0,988
10	0,175	0,558	0,935	0,993	1,000	1,000	1,000
25	0,345	0,892	1,000	1,000	1,000	1,000	1,000
50	0,552	0,973	1,000	1,000	1,000	1,000	1,000
100	0,756	0,994	1,000	1,000	1,000	1,000	1,000
200	0,869	1,000	1,000	1,000	1,000	1,000	1,000
300	0,918	1,000	1,000	1,000	1,000	1,000	1,000

On remarque très clairement que le test de Frees est plus performant dans ce contexte. Étant donné qu'il se base sur une moyenne des carrés des corrélations, le test est plus puissant pour une corrélation qui peut alterner de signe. Le test de

Pesaran donne de meilleurs résultats que celui de Friedman. Cependant, les deux tests présentent une puissance qui ne varie qu'avec T . En effet, comme ces deux tests se basent sur une moyenne des corrélations sans valeur absolue ni mise au carré, ajouter des observations ne permet pas une meilleure puissance puisque des corrélations positives s'annulent avec des corrélations négatives. Par contre plus T est grand plus le calcul de la corrélation sera précis et les tests performants.

2.3. RÉSULTATS THÉORIQUES CONCERNANT LES MODÈLES

Pesaran (2004) ne mentionne pas la covariance théorique entre les termes d'erreurs de son modèle. On dérive ici la forme théorique des covariances utilisées dans les modèles des deux sections précédentes. On commence par la covariance entre les différents termes d'erreurs u_{it} en fonction de la force de la dépendance choisie. On rappelle que ces termes sont les mêmes dans les deux modèles utilisés dans les simulations précédentes. Nous développons ces formules afin d'avoir plus d'informations sur les modèles utilisés dans ce mémoire et savoir comment les différents paramètres des modèles interviennent dans la calibration des différentes dépendances étudiées. On utilise aussi les différentes formules théoriques du modèle simplifié pour pouvoir approximer celles du modèle linéaire généralisé.

Dans la suite, on utilise à plusieurs reprises la formule suivante pour trouver les différents moments d'une variable aléatoire uniforme X sur l'intervalle (θ_0, θ_1) . Ainsi,

$$E(X^k) = \frac{\theta_1^{k+1} - \theta_0^{k+1}}{(k+1)(\theta_1 - \theta_0)}.$$

Démonstration :

En se rappelant que la densité d'une loi uniforme est $f(x) = \frac{1}{\theta_1 - \theta_0}$ pour $\theta_0 < x < \theta_1$:

$$E(X^k) = \int_{\theta_0}^{\theta_1} \frac{x^k}{\theta_1 - \theta_0} dx = \left[\frac{x^{k+1}}{(k+1)(\theta_1 - \theta_0)} \right]_{\theta_0}^{\theta_1} = \frac{\theta_1^{k+1} - \theta_0^{k+1}}{(k+1)(\theta_1 - \theta_0)}.$$

En considérant les variables permettant de générer les simulations comme fixes, seuls les variables ϵ_{it} et f_t sont considérés comme aléatoires lors de l'inférence. Les autres variables sont fixées au moment de faire les tests. Dans ce contexte on s'intéresse donc à la covariance entre les termes d'erreurs conditionnellement aux paramètres utilisés pour générer la simulation, c'est à dire : $Cov(u_{it}, u_{jt} | \gamma_i, \gamma_j)$.

Proposition 2.3.1. *La covariance entre les deux erreurs de deux individus différents ($i \neq j$) conditionnellement aux paramètres des simulations, au même temps t , est la suivante :*

$$Cov(u_{it}, u_{jt} | \gamma_i, \gamma_j) = \gamma_i \gamma_j.$$

De plus dans le modèle de Pesaran la covariance entre deux individus est (en notant $\Theta = (\mu_i, \beta_i, \gamma_i, \mu_j, \beta_j, \gamma_j)$) :

$$Cov(y_{it}^A, y_{jt}^A | \Theta) = \sum_{k=0}^{t-1} \beta_i^k \beta_j^k \gamma_i \gamma_j,$$

et dans le modèle simplifié (en notant $\Theta = (\alpha_i, \beta_i, \gamma_i, \alpha_j, \beta_j, \gamma_j)$) :

$$Cov(y_{it}^L, y_{jt}^L | \Theta) = \gamma_i \gamma_j.$$

Démonstration : On a tout d'abord :

$$Cov(u_{it}, u_{jt} | \gamma_i, \gamma_j) = Cov(\epsilon_{it} + \gamma_i f_t, \epsilon_{jt} + \gamma_j f_t | \gamma_i, \gamma_j).$$

Comme les ϵ_{it} sont indépendants il va rester directement :

$$\begin{aligned}
Cov(u_{it}, u_{jt} | \gamma_i, \gamma_j) &= Cov(\gamma_i f_t, \gamma_j f_t | \gamma_i, \gamma_j) \\
&= \gamma_i \gamma_j Cov(f_t, f_t) \\
&= \gamma_i \gamma_j V(f_t) \\
&= \gamma_i \gamma_j.
\end{aligned}$$

Il serait assez simple de montrer que c'est également cette covariance entre les individus dans le cas du modèle simplifié. Pour le modèle de Pesaran le terme autorégressif fait apparaître une complication :

$$\begin{aligned}
Cov(y_{it}^A, y_{jt}^A | \Theta) &= Cov(\mu_i(1 - \beta_i) + \beta_i y_{i(t-1)}^A + u_{it}, \mu_j(1 - \beta_j) + \beta_j y_{j(t-1)}^A + u_{jt} | \Theta) \\
&= Cov(\beta_j y_{i(t-1)}^A, \beta_j y_{j(t-1)}^A | \Theta) + Cov(u_{it}, u_{jt} | \Theta) \\
&= \beta_i \beta_j Cov(y_{i(t-1)}^A, y_{j(t-1)}^A | \Theta) + \gamma_i \gamma_j.
\end{aligned}$$

Cette forme recursive donne la formule voulue car $y_{i0}^A = 0$.

Nous nous intéressons dans la suite aux différentes formules théoriques en supposant l'ensemble des variables utilisées comme aléatoires.

Proposition 2.3.2. *La covariance entre les deux erreurs de deux individus différents ($i \neq j$) au même temps t , en notant a et b tels que $\gamma_i \sim \text{Siid } U(a; b)$, est la suivante :*

$$Cov(u_{it}, u_{jt}) = \left(\frac{a+b}{2} \right)^2.$$

Démonstration :

$$Cov(u_{it}, u_{jt}) = Cov(\epsilon_{it} + \gamma_i f_t, \epsilon_{jt} + \gamma_j f_t).$$

Par indépendance des ϵ_{it} , on a donc :

$$\begin{aligned}
Cov(u_{it}, u_{jt}) &= Cov(\gamma_i f_t, \gamma_j f_t) \\
&= E(\gamma_i f_t \gamma_j f_t) - E(\gamma_i f_t) E(\gamma_j f_t) \\
&= E(\gamma_i) E(\gamma_j) E(f_t^2) - E(\gamma_i) E(\gamma_j) E(f_t)^2 \\
&= E(\gamma_i) E(\gamma_j) V(f_t).
\end{aligned}$$

On rappelle que f_t suit une loi normale standard donc sa variance $V(f_t) = 1$ et comme γ_i suit une loi uniforme, en notant les deux bornes de la distribution uniforme a et b on obtient la formule proposée.

Proposition 2.3.3. *La covariance entre les deux erreurs d'un même individu à deux temps différents ($t \neq s$) est la suivante :*

$$Cov(u_{it}, u_{is}) = 0.$$

Démonstration :

$$Cov(u_{it}, u_{is}) = Cov(\epsilon_{it} + \gamma_i f_t, \epsilon_{is} + \gamma_i f_s).$$

Comme ci-dessus, avec l'indépendance des ϵ_{it} , on a donc :

$$\begin{aligned}
Cov(u_{it}, u_{is}) &= Cov(\gamma_i f_t, \gamma_i f_s) \\
&= E(\gamma_i f_t \gamma_i f_s) - E(\gamma_i f_t) E(\gamma_i f_s) \\
&= E(\gamma_i^2) E(f_t) E(f_s) - E(\gamma_i)^2 E(f_t) E(f_s) \\
&= E(f_t) E(f_s) V(\gamma_i).
\end{aligned}$$

Comme γ_i est uniforme $V(\gamma_i) = \frac{b^2 - a^2}{12}$ et comme $E(f_t) = 0$ on a finalement ce que l'on cherchait.

Enfin, il serait facile de montrer que $Cov(u_{it}, u_{js})$ est également nul pour tout $i \neq j$ et $t \neq s$.

On voit qu'il est ici assez simple de contrôler la covariance entre les erreurs uniquement d'un individu à l'autre grâce au terme $\gamma_i f_t$. On peut aussi trouver la variance des erreurs :

$$\begin{aligned}
V(u_{it}) &= V(\epsilon_{it} + \gamma_i f_t) \\
&= V(\epsilon_{it}) + V(\gamma_i f_t) \\
&= 1 + E(\gamma_i^2 f_t^2) - E(\gamma_i)^2 E(f_t)^2 \\
&= 1 + E(\gamma_i^2) E(f_t^2) - 0 \\
&= 1 + \frac{b^3 - a^3}{3(b-a)} \text{ car } \gamma_i \text{ est uniforme} \\
&= 1 + \frac{b^2 + ab + a^2}{3}.
\end{aligned}$$

On va s'intéresser maintenant à la covariance entre les termes y_{it} dans les deux modèles. Commençons par le modèle de Pesaran (2004).

Proposition 2.3.4. *La covariance entre deux éléments dans le modèle de Pesaran (2004) à un même temps et individus différents ($i \neq j$) est :*

$$Cov(y_{it}^A, y_{jt}^A) = \left(\frac{a+b}{2}\right)^2 \left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \dots + \frac{1}{t^2}\right).$$

Démonstration :

$$Cov(y_{it}^A, y_{jt}^A) = Cov(\mu_i(1 - \beta_i) + \beta_i y_{i(t-1)}^A + u_{it}, \mu_j(1 - \beta_j) + \beta_j y_{j(t-1)}^A + u_{jt}).$$

Par indépendance des μ_i et β_i (entre eux et sur certains autres termes), on peut donc réduire à la forme suivante :

$$\begin{aligned} Cov(y_{it}^A, y_{jt}^A) &= Cov(\beta_i y_{i(t-1)}^A + u_{it}, \beta_j y_{j(t-1)}^A + u_{jt}) \\ &= Cov(\beta_i y_{i(t-1)}^A, \beta_j y_{j(t-1)}^A) + Cov(u_{it}, u_{jt}). \end{aligned}$$

Le deuxième terme de la covariance est connu d'après la proposition 2.3.2. Il faut donc trouver le premier terme. Pour cela on utilise la formule de la covariance conditionnelle :

$$Cov(X, Y) = E(Cov(X, Y|Z)) + Cov(E(X|Z), E(Y|Z)).$$

En prenant ici $Z = (\beta_i, \beta_j)$ on a alors :

$$\begin{aligned} Cov(\beta_i y_{i(t-1)}^A, \beta_j y_{j(t-1)}^A) &= E(Cov(\beta_i y_{i(t-1)}^A, \beta_j y_{j(t-1)}^A | \beta_i, \beta_j)) \\ &\quad + Cov(E(\beta_i y_{i(t-1)}^A | \beta_i, \beta_j), E(\beta_j y_{j(t-1)}^A | \beta_i, \beta_j)) \\ &= E(\beta_i \beta_j Cov(y_{i(t-1)}^A, y_{j(t-1)}^A | \beta_i, \beta_j)). \end{aligned}$$

En effet, le second terme $Cov(E(\beta_i y_{i(t-1)}^A | \beta_i, \beta_j), E(\beta_j y_{j(t-1)}^A | \beta_i, \beta_j))$ est nul puisque le terme $E(\beta_i y_{i(t-1)}^A | \beta_i, \beta_j)$ sera une fonction uniquement de β_i et β_j . Comme β_j n'apparaît pas dans le développement de $y_{i(t-1)}^A$, ce terme sera donc fonction uniquement de β_i . De même le terme $E(\beta_j y_{j(t-1)}^A | \beta_i, \beta_j)$ ne sera fonction que de β_j . Ces deux variables étant indépendantes (pour $i \neq j$), la covariance de deux fonctions de ces variables sera nulle. Il reste donc à trouver :

$$E(\beta_i \beta_j Cov(y_{i(t-1)}^A, y_{j(t-1)}^A | \beta_i, \beta_j)).$$

Comme on peut le voir, cela fait intervenir le terme $Cov(y_{i(t-1)}^A, y_{j(t-1)}^A | \beta_i, \beta_j)$ qui en suivant les mêmes étapes donne :

$$Cov(y_{i(t-1)}^A, y_{j(t-1)}^A | \beta_i, \beta_j) = Cov(\beta_i y_{i(t-2)}^A, \beta_j y_{j(t-2)}^A | \beta_i, \beta_j)$$

$$\begin{aligned}
& + Cov(u_{i(t-1)}, u_{j(t-1)} | \beta_i, \beta_j) \\
& = \left(\frac{a+b}{2} \right)^2 + \beta_i \beta_j Cov(y_{i(t-2)}^A, y_{j(t-2)}^A | \beta_i, \beta_j).
\end{aligned}$$

On a alors finalement :

$$Cov(y_{it}^A, y_{jt}^A) = \left(\frac{a+b}{2} \right)^2 + E \left(\beta_i \beta_j \left(\frac{a+b}{2} \right)^2 + \beta_i^2 \beta_j^2 Cov(y_{i(t-2)}^A, y_{j(t-2)}^A | \beta_i, \beta_j) \right)$$

C'est une forme réursive sur t . Sachant que $Cov(y_{i0}^A, y_{j0}^A) = 0$, car les deux termes sont fixes (ou indépendants, générés aléatoirement), on a alors :

$$\begin{aligned}
Cov(y_{it}^A, y_{jt}^A) & = \left(\frac{a+b}{2} \right)^2 \left(1 + E(\beta_i \beta_j) + E(\beta_i^2 \beta_j^2) + \dots + E(\beta_i^{t-1} \beta_j^{t-1}) \right) \\
& = \left(\frac{a+b}{2} \right)^2 \left(1 + E(\beta_i)E(\beta_j) + E(\beta_i^2)E(\beta_j^2) + \dots + E(\beta_i^{t-1})E(\beta_j^{t-1}) \right) \\
& = \left(\frac{a+b}{2} \right)^2 \left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \dots + \frac{1}{t^2} \right).
\end{aligned}$$

On a utilisé ici la formule des moments pour β_i qui est uniforme ($\theta_0 = 0, \theta_1 = 1$). Il est intéressant de noter que la covariance tend donc rapidement vers $\left(\frac{a+b}{2} \right)^2 \frac{\pi^2}{6}$ lorsque t augmente. Cette convergence très connue appelée problème de Bâle est résolue par Euler et démontrée par exemple dans Harper (2003).

Proposition 2.3.5. *Dans le cas où $i = j$:*

$$V(y_{it}^A) = 2 \left(1 - \frac{2}{t+1} + \frac{1}{2t+1} \right) + \sum_{k=0}^{t-1} \left(\frac{1}{2k+1} \left[1 + \frac{b^2 + ab + a^2}{3} \right] \right).$$

Démonstration : On peut tout d'abord voir le développement suivant :

$$\begin{aligned}
y_{it}^A & = \mu_i(1 - \beta_i) + \beta_i y_{i(t-1)}^A + u_{it} \\
& = \mu_i(1 - \beta_i) + \beta_i [\mu_i(1 - \beta_i) + \beta_i y_{i(t-2)}^A + u_{i(t-1)}] + u_{it}
\end{aligned}$$

$$\begin{aligned}
&= \mu_i(1 - \beta_i)(1 + \beta_i + \beta_i^2 + \dots + \beta_i^{t-1}) + u_{it} + \beta_i u_{i(t-1)} + \dots + u_{i1} \beta_i^{t-1} \\
&= \mu_i(1 - \beta_i^t) + u_{it} + \beta_i u_{i(t-1)} + \dots + u_{i1} \beta_i^{t-1}.
\end{aligned}$$

On remarquera les résultats suivants :

- $E(u_{it}) = E(\epsilon_{it}) + E(\gamma_i)E(f_t) = 0$.
- $\mu_i \sim N(0;2)$.
- $Cov(u_{it}, u_{is}) = 0$ si $t \neq s$.
- On notera par la suite $A_{it} = \mu_i(1 - \beta_i^t)$ et donc $E(A_{it}) = E(\mu_i)E(1 - \beta_i^t) = 0$.

On peut alors montrer :

$$\begin{aligned}
Cov(A_{it}, \beta_i^k u_{i(t-k)}) &= E(A_{it} \beta_i^k u_{i(t-k)}) - E(A_{it})E(\beta_i^k u_{i(t-k)}) \\
&= E(A_{it} \beta_i^k)E(u_{i(t-k)}) - E(A_{it})E(\beta_i^k u_{i(t-k)}) \\
&= 0.
\end{aligned}$$

Si $k \neq l$, en notant que $Cov(u_{it}, u_{is}) = 0$ implique $E(u_{it}u_{is}) = E(u_{it})E(u_{is})$, on a :

$$\begin{aligned}
Cov(\beta_i^k u_{i(t-k)}, \beta_i^l u_{i(t-l)}) &= E(\beta_i^k u_{i(t-k)} \beta_i^l u_{i(t-l)}) - E(\beta_i^k u_{i(t-k)})E(\beta_i^l u_{i(t-l)}) \\
&= E(\beta_i^{k+l})E(u_{i(t-k)})E(u_{i(t-l)}) - E(\beta_i^k)E(u_{i(t-k)})E(\beta_i^l)E(u_{i(t-l)}) \\
&= 0.
\end{aligned}$$

Ainsi, tous les termes de covariances dans $V(y_{it}^A)$ sont nuls et il reste uniquement :

$$V(y_{it}^A) = V(A_{it}) + V(u_{it}) + \dots + V(u_{i1} \beta_i^{t-1}).$$

Il nous reste à trouver :

$$V(A_{it}) = E(V(A_{it}|\beta_i)) + V(E(A_{it}|\beta_i))$$

$$\begin{aligned}
&= E(V(\mu_i(1 - \beta_i^t)|\beta_i)) + V(E(\mu_i(1 - \beta_i^t)|\beta_i)) \\
&= E(2(1 - \beta_i^t)^2) + 0 \\
&= E(2(1 - 2\beta_i^t + \beta_i^{2t})) \\
&= 2\left(1 - \frac{2}{t+1} + \frac{1}{2t+1}\right).
\end{aligned}$$

et

$$\begin{aligned}
V(\beta_i^k u_{i(t-k)}) &= E(\beta_i^{2k} u_{i(t-k)}^2) - E(\beta_i^k u_{i(t-k)})^2 \\
&= E(\beta_i^{2k})E(u_{i(t-k)}^2) - E(\beta_i^k)^2 E(u_{i(t-k)})^2 \\
&= E(\beta_i^{2k})V(u_{i(t-k)}) \\
&= \frac{1}{2k+1} \left(1 + \frac{b^2 + ab + a^2}{3}\right).
\end{aligned}$$

Ce qui donne bien la forme recherchée.

Proposition 2.3.6. *La covariance entre deux éléments dans le modèle de Pesaran (2004) d'un même individu à deux temps consécutifs est :*

$$Cov(y_{it}^A, y_{i(t-1)}^A) = 2\left(1 - \frac{3t+1}{2t(t+1)}\right) + \sum_{k=1}^{t-1} \left(\frac{1}{2k} \left[1 + \frac{b^2 + ab + a^2}{3}\right]\right).$$

Démonstration :

En reprenant $y_{it}^A = A_{it} + u_{it} + \dots + u_{i1}\beta_i^{t-1}$ on a directement :

$$Cov(y_{it}^A, y_{i(t-1)}^A) = Cov(A_{it} + u_{it} + \dots + u_{i1}\beta_i^{t-1}, A_{i(t-1)} + u_{i(t-1)} + \dots + u_{i1}\beta_i^{t-2}).$$

On a vu précédemment que $Cov(A_{it}, \beta_i^k u_{i(t-k)}) = 0$, une démonstration similaire donne directement $Cov(A_{it}, \beta_i^k u_{is}) = 0$ pour tout s . On a également vu

que si $k \neq l$ les termes $Cov(\beta_i^k u_{i(t-k)}, \beta_i^l u_{i(t-l)}) = 0$ et de même un développement similaire donnerait $Cov(\beta_i^k u_{is}, \beta_i^l u_{ip}) = 0$ avec $s \neq p$. Avec tous ces termes de covariances nuls il restera alors :

$$\begin{aligned} Cov(y_{it}^A, y_{i(t-1)}^A) &= Cov(A_{it}, A_{i(t-1)}) + Cov(u_{i(t-1)}, \beta_i u_{i(t-1)}) \\ &\quad + Cov(\beta_i u_{i(t-2)}, \beta_i^2 u_{i(t-2)}) + \dots + Cov(\beta_i^{t-1} u_{i1}, \beta_i^{t-2} u_{i1}). \end{aligned}$$

Il nous reste donc à trouver :

$$\begin{aligned} Cov(A_{it}, A_{i(t-1)}) &= Cov(\mu_i(1 - \beta_i^t), \mu_i(1 - \beta_i^{t-1})) \\ &= E(\mu_i^2(1 - \beta_i^t)(1 - \beta_i^{t-1})) - E(\mu_i)E(1 - \beta_i^{t-1})E(1 - \beta_i^t) \\ &= V(\mu_i)E(1 - \beta_i^t - \beta_i^{t-1} + \beta_i^{2t-1}) - 0 \\ &= 2 \left(1 - \frac{1}{t+1} - \frac{1}{t} + \frac{1}{2t} \right) \\ &= 2 \left(1 - \frac{3t+1}{2t(t+1)} \right), \end{aligned}$$

et

$$\begin{aligned} Cov(\beta_i^{k-1} u_{i(t-k)}, \beta_i^k u_{i(t-k)}) &= E(\beta_i^{k-1} u_{i(t-k)} \beta_i^k u_{i(t-k)}) - E(\beta_i^{k-1} u_{i(t-k)})E(\beta_i^k u_{i(t-k)}) \\ &= E(\beta_i^{2k-1})E(u_{i(t-k)}^2) - E(\beta_i^{k-1})E(u_{i(t-k)})E(\beta_i^k)E(u_{i(t-k)}) \\ &= E(\beta_i^{2k-1})V(u_{i(t-k)}) \\ &= \frac{1}{2k} V(u_{i(t-k)}) \\ &= \frac{1}{2k} \left(1 + \frac{b^2 + ab + a^2}{3} \right). \end{aligned}$$

En remplaçant dans $Cov(y_{it}^A, y_{i(t-1)}^A)$ on aura ce qu'on voulait.

Dans le modèle simplifié (équation (2.1.3)) il est assez simple de voir que puisque les variables α_i , β_i et les x_{it} sont toutes indépendantes on aura directement les résultats suivants.

Proposition 2.3.7. *Dans le modèle simplifié, la covariance entre deux individus à un même temps t est :*

$$\begin{aligned} Cov(y_{it}^L, y_{jt}^L) &= Cov(\alpha_i + \beta_i x_{it} + u_{it}, \alpha_j + \beta_j x_{jt} + u_{jt}) \\ &= Cov(u_{it}, u_{jt}) \\ &= \left(\frac{a+b}{2}\right)^2. \end{aligned}$$

Dans ce cas-ci on peut également facilement calculer $V(y_{it})$. On aura :

$$\begin{aligned} V(y_{it}^L) &= V(\alpha_i) + V(\beta_i x_{it}) + V(u_{it}) \\ &= \frac{1}{12} + E(\beta_i^2 x_{it}^2) - [E(\beta_i x_{it})]^2 + 1 + \frac{b^3 - a^3}{3(b-a)} \\ &= \frac{13}{12} + \frac{5}{3} - 1 + \frac{b^2 + ab + a^2}{3} \\ &= \frac{7}{4} + \frac{b^2 + ab + a^2}{3}. \end{aligned}$$

Proposition 2.3.8. *Dans le modèle simplifié :*

$$Cov(y_{it}^L, y_{i(t-1)}^L) = \frac{5}{12}.$$

Démonstration :

$$\begin{aligned}
Cov(y_{it}^L, y_{i(t-1)}^L) &= Cov(\alpha_i + \beta_i x_{it} + u_{it}, \alpha_i + \beta_i x_{i(t-1)} + u_{i(t-1)}) \\
&= Cov(\alpha_i, \alpha_i) + Cov(\beta_i x_{it}, \beta_i x_{i(t-1)}) + Cov(u_{it}, u_{i(t-1)}) \\
&= \frac{1}{12} + E(\beta_i x_{it} \beta_i x_{i(t-1)}) - E(\beta_i x_{it}) E(\beta_i x_{i(t-1)}) + 0 \\
&= \frac{1}{12} + E(x_{it}) E(x_{i(t-1)}) E(\beta_i^2) - E(x_{it}) E(x_{i(t-1)}) E(\beta_i)^2 \\
&= \frac{1}{12} + E(x_{it}) E(x_{i(t-1)}) V(\beta_i) \\
&= \frac{1}{12} + 4 \times \frac{1}{12} \\
&= \frac{5}{12}.
\end{aligned}$$

On peut alors vérifier que l'ensemble des covariances ici trouvées rentrent bien dans le cadre des présupposés qui avaient été présentés dans Pesaran (2004) à la section 1.1.1. Principalement le fait que les erreurs sont bien indépendantes dans le temps et ont de la dépendance uniquement entre les individus.

Chapitre 3

SIMULATIONS POUR UN MODÈLE LINÉAIRE GÉNÉRALISÉ

Dans ce chapitre, nous exposons notre modèle pour un modèle linéaire généralisé (GLM) de Poisson. En s'inspirant de Hsiao *et al.* (2007), nous appliquons les trois différents tests à ce modèle. Dans cet article, ils utilisent des GLM probit et tobit pour des données longitudinales et appliquent la statistique du *CD* de Pesaran. Nous proposons donc ici d'appliquer les tests de Pesaran, de Friedman et de Frees dans le cadre d'un modèle linéaire généralisé de Poisson. Pour plus de détails sur les GLM, voir par exemple le texte de référence McCullagh et Nelder (1989) au chapitre 2 ainsi que Dobson (2002) au chapitre 3 dont nous reprenons l'introduction aux GLM ici. Jusqu'à présent, dans les deux premiers chapitres, nous avons considéré des modèles linéaires de la forme $E(Y|X) = X\beta$. Les modèles linéaires généralisés permettent d'avoir une fonction autre que l'identité entre l'espérance de la réponse Y et les variables explicatives X et d'avoir un éventail plus large de distributions pour Y . En d'autres termes nous pouvons relier la réponse et ses variables explicatives par une relation du type :

$$E(Y|X) = h(X\beta),$$

où β serait un vecteur de paramètres et X une matrice des variables explicatives. Une manière alternative dans la littérature est d'écrire $g(E(Y_i|X)) = g(\mu_i) = X_i\beta$.

Dans cette dernière notation la fonction g est appelée fonction de lien. La variable Y peut ensuite suivre une des distributions de la famille de lois exponentielles. Une variable fait partie de cette famille lorsque sa fonction de répartition en fonction de ses paramètres θ est de la forme

$$f(y; \theta) = \exp(a(y)b(\theta) + c(\theta) + d(y))$$

avec les fonctions a , b , c , d variant d'une distribution à l'autre. La fonction b est parfois appelé le paramètre naturel. Dans notre contexte ici, nous nous intéressons au modèle d'une régression de Poisson. Dans ce cas, la fonction de masse d'une variable de Poisson de paramètre λ en y entier positif est :

$$f(y; \lambda) = \frac{\lambda^y e^{-\lambda}}{y!}.$$

On reconnaîtra donc que cette distribution fait partie de la famille exponentielle avec $b(\lambda) = \ln(\lambda)$, $c(\lambda) = -\lambda$ et $d = -\ln(y!)$ et en voyant que

$$f(y; \lambda) = \exp(y \ln(\lambda) - \lambda - \ln(y!)).$$

Cette distribution est en général utilisée pour dénombrer des évènements indépendants dans une période ou un espace. La forme générale d'un modèle linéaire généralisé de Poisson sera alors telle que $Y|X$ suit une loi de Poisson et $E(Y|X) = \exp(X\beta)$. En notant $E(Y|X) = \lambda$ le paramètre de la loi de Poisson, on a alors $\ln(\lambda) = X\beta$ avec le logarithme naturel comme fonction de lien.

Dans la section suivante, nous introduisons précisément le modèle qui sera utilisé dans les simulations et auquel nous appliquerons les trois tests vus dans le chapitre un.

3.1. MODÈLE DE SIMULATION

Le modèle que nous allons considérer et évaluer lors des simulations sera le modèle suivant :

$$\log(\lambda_{it}) = \alpha_i + \beta_i x_{it}.$$

Sous ce modèle nous aurons alors des données de Poisson indépendantes. Cependant, afin de faire apparaître une dépendance entre les individus et lors de la génération de nos simulations, nous insérons un facteur commun f_t dans la moyenne. Pour les simulations qui suivront à la section suivante, nous choisissons donc d'utiliser le modèle de simulation suivant :

$$\log(\lambda_{it}) = \alpha_i + \beta_i x_{it} + \gamma_i f_t, \quad (3.1.1)$$

où comme dans le modèle linéaire simplifié $\alpha_i \sim \text{Siid } U(0;1)$, $\beta_i \sim \text{Siid } U(0;1)$ et $x_{it} \sim \text{Siid } N(2,1)$. La variable y_{it}^P est ensuite générée comme une variable de Poisson de paramètre λ_{it} . Il nous faut remarquer que dans le modèle de simulation présenté, y_{it}^P n'est pas une variable de Poisson. Seulement $y_{it}^P | \lambda_{it}$ est une variable de Poisson. En considérons que seul f_t est aléatoire, les autres variables étant seulement générées aléatoirement pour les données simulées, la distribution marginale serait :

$$P(y_{it} = y | \alpha_i, \beta_i, x_{it}, \gamma_i) = \int_{f_t} P(y_{it} = y | \lambda_{it}) g(f_t) df_t.$$

où g est la densité de la variable f_t . En effet, pour que la variable y_{it}^P soit Poisson, il faut que son paramètre soit fixe. Or dans le modèle choisi, λ_{it} n'est pas fixe afin de créer de la dépendance par l'intermédiaire de $\gamma_i f_t$. Dans le cas contraire si les paramètres étaient fixés alors les variables y_{it}^P ne pourraient pas être dépendantes avec un tel modèle. Par abus de langage, nous continuons de l'appeler modèle linéaire généralisé de Poisson dans la suite.

Dans la section 3.3 nous montrons que λ_{it} sera d'un ordre de grandeur assez grand afin que, pour un i donné, y_{it}^P ait une probabilité faible d'être nul pour tout t . En effet, si λ_{it} est trop petit, il y aurait un risque que les variables y_{it}^P simulées soient parfois toutes nulles pour un individu, voir pour tous les individus. Les calculs pour trouver une approximation de la valeur de $E(\lambda_{it})$ et justifiant ainsi les paramètres de simulation utilisés sont faits en section 3.3 en utilisant la méthode delta.

Une des difficultés d'un modèle linéaire généralisé est l'ambiguïté des résidus. Étant donné que nous réalisons par la suite des régressions de Poisson dans le cadre de cette étude, on utilise les résidus de Pearson. En effet, puisque la variance et l'espérance d'une Poisson sont identiques nous avons $V(y_{it}^P|x_{it},\hat{\theta}_i) = E(y_{it}^P|x_{it},\hat{\theta}_i)$ (avec $\hat{\theta}_i$ l'ensemble des paramètres estimés). Les résidus du type réponse $y_{it}^P - E(y_{it}^P|x_{it},\hat{\theta}_i)$ seront hétéroscédastiques. La littérature suggère donc, pour une régression de Poisson, d'utiliser soit les résidus de Pearson, soit ceux de la déviance (voir Fox et Sanford 2011 section 6.6). Nous utilisons ici ceux de Pearson. D'autres essais avec d'autres types de résidus sont faits brièvement en annexe. On voit que cela ne change pas nos résultats pour le modèle utilisé (résidus de type réponse ou déviance). On utilise donc ici les résidus suivants :

$$e_{it}^P = \frac{y_{it}^P - E(y_{it}^P|x_{it},\hat{\theta}_i)}{\sqrt{V(y_{it}^P|x_{it},\hat{\theta}_i)}}, \quad (3.1.2)$$

où $\hat{\theta}_i$ est l'estimateur des différents paramètres. Avec cette définition des résidus, on peut estimer la corrélation échantillonnale des résidus :

$$\hat{\rho}_{ij} = \frac{\sum_{t=1}^T (e_{it}^P - \bar{e}_i^P)(e_{jt}^P - \bar{e}_j^P)}{\left(\sum_{t=1}^T (e_{it}^P - \bar{e}_i^P)^2 \right)^{\frac{1}{2}} \left(\sum_{t=1}^T (e_{jt}^P - \bar{e}_j^P)^2 \right)^{\frac{1}{2}}}.$$

où \bar{e}_i^P représente la moyenne (échantillonnale) des e_{it}^P . Hsiao *et al.* (2007) utilisent ce procédé pour calculer le CD dans le cas d'un modèle tobit et d'un modèle probit. Leurs conclusions sont similaires au modèle linéaire. Ils font de plus la comparaison avec le CD_{lm} qui s'avère, comme pour Pesaran, présenter des problèmes dans certaines situations et concluent que le CD est un test plus pratique en général. Nous nous proposons ici d'utiliser également les tests de Friedman et Frees avec la corrélation de Spearman.

Nous exposons maintenant comme au chapitre précédent les différentes étapes des simulations utilisées ainsi que les résultats. Pour des paramètres N , T et γ_i donnés, une itération est générée de la manière suivante :

Étape 1 : Générer y_{it}^P

Les paramètres α_i et β_i sont générés par des $U(0;0,1)$ et $U(0;0,1)$ respectivement, et x_{it} par des $N(2;1)$ respectivement. La variable simulant la dépendance, γ_i , est toujours générée avec une certaine uniforme selon le degré de dépendance simulé (voir section 2.1.1). Nous testerons également les performances des tests avec $\gamma_i \sim Siid N(0;1)$. Les variables f_t sont distribuées selon des $N(0;0,1)$. On construit alors

$$\lambda_{it} = \exp(\alpha_i + \beta_i x_{it} + \gamma_i f_t).$$

Puis on génère y_{it}^P à partir d'une Poisson de paramètre λ_{it} de telle sorte que $y_{it}^P | \lambda_{it}$ est une variable de Poisson. Si pour i fixe, les y_{it}^P générés sont tous égaux (variance nulle), la ligne est alors supprimée (cela peut se produire pour de petits T comme $T = 5$). Cependant, nous nous sommes arrangés pour que λ_{it} soit assez grand afin que cela arrive très rarement. Nos simulations montrent que seulement $T = 5$ crée ce genre de scénario et nous supprimons environ une ligne sur 1000 dans ce cas, ce qui est négligeable (entre 0,09% et 0,12%).

Étape 2 : Faire les régressions

Pour chaque i , on réalise une régression de Poisson de y_{it}^P selon x_{it} avec une ordonnée à l'origine ($t = 1, \dots, T$). Les résidus de Pearson de cette régression, e_{it}^P , sont ensuite calculés ainsi que les corrélations deux à deux entre ces résidus. Nous calculons également les corrélations de Spearman correspondantes à partir des rangs des différents e_{it}^P .

Étape 3 : Calcul du CD et décision

Si $\hat{\rho}_{ij}$ est la corrélation entre les résidus de Pearson e_{it}^P et e_{jt}^P , alors on calcule :

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij} \right).$$

Si la valeur absolue du CD calculée est supérieure à 1,96, alors on détecte de la dépendance pour cette itération. Si $\hat{\rho}_{S,ij}$ est la corrélation de Spearman, alors nous calculons les deux choses suivantes :

$$R_M = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{S,ij}$$

et

$$R_M^2 = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{S,ij}^2.$$

À partir de ces quantités, on peut donc calculer les statistiques de Friedman et Frees qui sont respectivement :

$$FR = (T-1)[(N-1)R_M + 1]$$

et

$$FRE = N \left(R_M^2 - \frac{1}{T-1} \right).$$

Nous comparons donc la valeur FR au quantile 95% d'une chi-deux à $T-1$ degrés de liberté et rejetterons l'hypothèse nulle si FR est plus grand. Nous utilisons l'approximation suggérée par Frees (2004) en considérant que $\frac{FRE}{\sqrt{V(Q)}}$ converge

vers une loi normale standard. Donc en calculant

$$V(Q) = \frac{32}{25} \frac{(T+2)^2}{(T-1)^3(T+1)^2} + \frac{4}{25} \frac{(5T+6)^2(T-3)}{T(T-1)^2(T+1)^2},$$

nous avons $\frac{FRE}{\sqrt{V(Q)}}$ et rejetons si la valeur absolue est supérieure à 1,96.

Dans la section suivante, nous présentons les résultats de nos simulations avec le modèle présenté et les trois tests. Nous présentons les niveaux et les puissances selon les différentes distributions que peut prendre γ_i .

3.2. RÉSULTATS DES SIMULATIONS ET COMPARAISONS DES TESTS

Nous présentons ici les différents résultats de nos simulations pour le modèle linéaire généralisé présenté en section 3.1. Les simulations sont faites pour les trois tests avec 1000 itérations pour chaque combinaison de N , T et γ_i .

TABLE 3.1. Niveaux, modèle linéaire généralisé $\gamma_i = 0$, test du CD de Pesaran

N\T	5	10	25	50	100	200	300
5	0,060	0,050	0,051	0,053	0,055	0,059	0,054
10	0,069	0,037	0,042	0,043	0,037	0,054	0,042
25	0,067	0,045	0,054	0,048	0,050	0,055	0,042
50	0,066	0,043	0,033	0,052	0,060	0,043	0,059
100	0,053	0,051	0,046	0,056	0,045	0,044	0,056
200	0,074	0,058	0,059	0,054	0,045	0,046	0,045
300	0,074	0,053	0,056	0,049	0,059	0,054	0,043

TABLE 3.2. Niveaux, modèle linéaire généralisé $\gamma_i = 0$, test du FR de Friedman

N\T	5	10	25	50	100	200	300
5	0,037	0,034	0,030	0,033	0,039	0,040	0,036
10	0,052	0,030	0,035	0,040	0,039	0,046	0,038
25	0,050	0,053	0,049	0,044	0,047	0,055	0,051
50	0,053	0,051	0,046	0,056	0,045	0,056	0,046
100	0,046	0,051	0,044	0,056	0,044	0,047	0,053
200	0,050	0,064	0,056	0,050	0,051	0,053	0,041
300	0,057	0,051	0,052	0,045	0,053	0,066	0,041

TABLE 3.3. Niveaux, modèle linéaire généralisé $\gamma_i = 0$, test du FRE de Frees

N\T	5	10	25	50	100	200	300
5	0,062	0,075	0,063	0,059	0,045	0,048	0,066
10	0,058	0,059	0,055	0,064	0,046	0,062	0,054
25	0,033	0,049	0,047	0,046	0,053	0,058	0,062
50	0,050	0,040	0,042	0,044	0,039	0,046	0,056
100	0,045	0,033	0,058	0,058	0,045	0,053	0,046
200	0,040	0,042	0,044	0,066	0,048	0,057	0,046
300	0,042	0,039	0,052	0,044	0,043	0,043	0,051

Comme pour les simulations avec des modèles linéaires, le niveau attendu doit se situer entre 3,65% et 6,35%. Ceci est le cas pour la plupart des valeurs du tableau. On peut noter quelques écarts en particulier lorsque $T = 5$ pour le test

du CD . C'est le même défaut que nous avons relevé pour les modèles du chapitre deux. Les tableaux de puissance suivants sont présentés trois par trois pour une même force de dépendance mais pour les trois tests différents.

TABLE 3.4. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,1;0,3)$, test du CD de Pesaran

N\T	5	10	25	50	100	200	300
5	0,194	0,314	0,580	0,801	0,933	0,993	0,997
10	0,367	0,605	0,902	0,994	1,000	1,000	1,000
25	0,651	0,948	0,998	1,000	1,000	1,000	1,000
50	0,846	0,995	1,000	1,000	1,000	1,000	1,000
100	0,942	1,000	1,000	1,000	1,000	1,000	1,000
200	0,980	1,000	1,000	1,000	1,000	1,000	1,000
300	0,989	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 3.5. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,1;0,3)$, test du FR de Friedman

N\T	5	10	25	50	100	200	300
5	0,114	0,240	0,505	0,765	0,911	0,990	0,998
10	0,280	0,533	0,889	0,993	0,999	1,000	1,000
25	0,588	0,928	0,997	1,000	1,000	1,000	1,000
50	0,800	0,992	1,000	1,000	1,000	1,000	1,000
100	0,925	1,000	1,000	1,000	1,000	1,000	1,000
200	0,975	1,000	1,000	1,000	1,000	1,000	1,000
300	0,989	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 3.6. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,1;0,3)$, test du FRE de Frees

N\T	5	10	25	50	100	200	300
5	0,065	0,114	0,293	0,585	0,806	0,961	0,985
10	0,090	0,173	0,560	0,866	0,982	0,998	1,000
25	0,095	0,384	0,912	0,996	1,000	1,000	1,000
50	0,174	0,640	0,989	1,000	1,000	1,000	1,000
100	0,270	0,852	0,999	1,000	1,000	1,000	1,000
200	0,439	0,952	1,000	1,000	1,000	1,000	1,000
300	0,527	0,973	1,000	1,000	1,000	1,000	1,000

TABLE 3.7. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du *CD* de Pesaran

N\T	5	10	25	50	100	200	300
5	0,328	0,538	0,907	0,984	1,000	1,000	1,000
10	0,566	0,874	0,999	1,000	1,000	1,000	1,000
25	0,860	0,995	1,000	1,000	1,000	1,000	1,000
50	0,950	1,000	1,000	1,000	1,000	1,000	1,000
100	0,989	1,000	1,000	1,000	1,000	1,000	1,000
200	0,996	1,000	1,000	1,000	1,000	1,000	1,000
300	0,997	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 3.8. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du *FR* de Friedman

N\T	5	10	25	50	100	200	300
5	0,194	0,459	0,870	0,976	0,998	1,000	1,000
10	0,473	0,846	0,996	0,999	1,000	1,000	1,000
25	0,804	0,990	1,000	1,000	1,000	1,000	1,000
50	0,926	1,000	1,000	1,000	1,000	1,000	1,000
100	0,982	1,000	1,000	1,000	1,000	1,000	1,000
200	0,995	1,000	1,000	1,000	1,000	1,000	1,000
300	0,995	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 3.9. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,2;0,4)$, test du *FRE* de Frees

N\T	5	10	25	50	100	200	300
5	0,101	0,233	0,681	0,931	0,994	1,000	1,000
10	0,130	0,455	0,901	0,996	1,000	1,000	1,000
25	0,244	0,776	0,998	1,000	1,000	1,000	1,000
50	0,400	0,929	0,999	1,000	1,000	1,000	1,000
100	0,558	0,985	1,000	1,000	1,000	1,000	1,000
200	0,744	0,997	1,000	1,000	1,000	1,000	1,000
300	0,777	0,997	1,000	1,000	1,000	1,000	1,000

TABLE 3.10. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du *CD* de Pesaran

N\T	5	10	25	50	100	200	300
5	0,467	0,786	0,991	1,000	1,000	1,000	1,000
10	0,747	0,967	1,000	1,000	1,000	1,000	1,000
25	0,937	0,999	1,000	1,000	1,000	1,000	1,000
50	0,978	1,000	1,000	1,000	1,000	1,000	1,000
100	0,996	1,000	1,000	1,000	1,000	1,000	1,000
200	1,000	1,000	1,000	1,000	1,000	1,000	1,000
300	0,999	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 3.11. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du *FR* de Friedman

N\T	5	10	25	50	100	200	300
5	0,334	0,689	0,983	1,000	1,000	1,000	1,000
10	0,658	0,950	1,000	1,000	1,000	1,000	1,000
25	0,909	0,998	1,000	1,000	1,000	1,000	1,000
50	0,971	1,000	1,000	1,000	1,000	1,000	1,000
100	0,994	1,000	1,000	1,000	1,000	1,000	1,000
200	0,998	1,000	1,000	1,000	1,000	1,000	1,000
300	0,999	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 3.12. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } U(0,3;0,5)$, test du *FRE* de Frees

N\T	5	10	25	50	100	200	300
5	0,147	0,433	0,929	0,994	1,000	1,000	1,000
10	0,206	0,722	0,995	1,000	1,000	1,000	1,000
25	0,408	0,937	1,000	1,000	1,000	1,000	1,000
50	0,606	0,992	1,000	1,000	1,000	1,000	1,000
100	0,803	1,000	1,000	1,000	1,000	1,000	1,000
200	0,884	0,999	1,000	1,000	1,000	1,000	1,000
300	0,912	1,000	1,000	1,000	1,000	1,000	1,000

Les résultats sur la puissance sont très similaires à ceux obtenus au chapitre deux lors des tests sur les modèles linéaires. Le *CD* présente les meilleurs résultats suivi par le test de Friedman. Le test de Frees présente une puissance moindre en général et surtout lorsque T est faible. Comme dans la section 2.2, le test de Frees pourrait présenter un avantage lorsque γ_i est centré en 0. Dans la suite nous réalisons donc une série de trois tableaux de puissance pour les trois tests en prenant $\gamma_i \sim N(0;1)$.

TABLE 3.13. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } N(0;1)$, test du *CD* de Pesaran

N\T	5	10	25	50	100	200	300
5	0,081	0,151	0,267	0,397	0,516	0,645	0,700
10	0,086	0,216	0,487	0,707	0,857	0,913	0,947
25	0,121	0,436	0,920	0,995	1,000	1,000	1,000
50	0,179	0,703	0,993	1,000	1,000	1,000	1,000
100	0,296	0,921	1,000	1,000	1,000	1,000	1,000
200	0,507	0,984	1,000	1,000	1,000	1,000	1,000
300	0,652	0,998	1,000	1,000	1,000	1,000	1,000

TABLE 3.14. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } N(0;1)$, test du *FR* de Friedman

N\T	5	10	25	50	100	200	300
5	0,047	0,085	0,160	0,213	0,256	0,296	0,293
10	0,056	0,103	0,185	0,220	0,272	0,269	0,301
25	0,065	0,123	0,199	0,266	0,320	0,339	0,361
50	0,092	0,134	0,287	0,356	0,453	0,533	0,601
100	0,119	0,245	0,524	0,711	0,885	0,975	0,986
200	0,167	0,544	0,915	0,988	1,000	1,000	1,000
300	0,267	0,744	0,988	1,000	1,000	1,000	1,000

TABLE 3.15. Puissance, modèle linéaire généralisé avec dépendance $\gamma_i \sim \text{Siid } N(0;1)$, test du *FRE* de Frees

N\T	5	10	25	50	100	200	300
5	0,185	0,625	0,948	0,991	0,996	0,999	0,998
10	0,316	0,922	0,998	1,000	1,000	1,000	1,000
25	0,659	0,997	1,000	1,000	1,000	1,000	1,000
50	0,869	1,000	1,000	1,000	1,000	1,000	1,000
100	0,969	1,000	1,000	1,000	1,000	1,000	1,000
200	0,984	1,000	1,000	1,000	1,000	1,000	1,000
300	0,993	1,000	1,000	1,000	1,000	1,000	1,000

On remarque très clairement qu'ici le test de Frees est plus performant. Étant donné qu'il se base sur une moyenne des carrés des corrélations il rendra mieux compte d'une corrélation qui peut alterner de signe. Le test de Pesaran donne également de meilleurs résultats que celui de Friedman. Contrairement à la forme linéaire cependant, ces deux tests augmentent à la fois avec N et avec T (la puissance augmentait avec T uniquement dans le cas linéaire).

Dans la section qui suit, nous développons des approximations pour les valeurs moyennes de la variable λ_{it} et la covariance entre y_{it}^P et y_{jt}^P . Nous utilisons la méthode delta pour obtenir ces approximations.

3.3. RÉSULTATS THÉORIQUES CONCERNANT LE MODÈLE

Comme au début de la section 2.3 nous cherchons à savoir quelle serait la covariance entre deux individus dans ce modèle. Nous allons donc considérer les paramètres générés pour la simulation comme fixe car leur distribution est utilisée uniquement dans le cadre des simulations et n'est pas une hypothèse faite sur le modèle. Nous allons donc nous intéresser à ($i \neq j$) :

$$Cov(y_{it}^P, y_{jt}^P | \Theta),$$

où $\Theta = (\alpha_i, \beta_i, x_{it}, \alpha_j, \beta_j, x_{jt})$.

Ensuite, il suffit de constater que :

$$\begin{aligned} Cov(y_{it}^P, y_{jt}^P) &= Cov(E(y_{it}^P | \lambda_{it}, \lambda_{jt}), E(y_{jt}^P | \lambda_{it}, \lambda_{jt})) + E(Cov(y_{it}^P, y_{jt}^P | \lambda_{it}, \lambda_{jt})) \\ &= Cov(E(y_{it}^P | \lambda_{it}, \lambda_{jt}), E(y_{jt}^P | \lambda_{it}, \lambda_{jt})) \\ &= Cov(\lambda_{it}, \lambda_{jt}). \end{aligned}$$

Donc ici on aura :

$$\begin{aligned} Cov(y_{it}^P, y_{jt}^P | \Theta) &= Cov(\lambda_{it}, \lambda_{jt} | \Theta) \\ &= Cov(e^{\alpha_i + \beta_i x_{it} + \gamma_i f_t}, e^{\alpha_j + \beta_j x_{jt} + \gamma_j f_t} | \Theta) \\ &= e^{\alpha_i + \alpha_j + \beta_i x_{it} + \beta_j x_{jt}} Cov(e^{\gamma_i f_t}, e^{\gamma_j f_t} | \Theta) \\ &= C \left[E(e^{\gamma_i f_t} e^{\gamma_j f_t} | \Theta) - E(e^{\gamma_i f_t} | \Theta) E(e^{\gamma_j f_t} | \Theta) \right] \\ &= C \left[E(e^{(\gamma_i + \gamma_j) f_t} | \Theta) - E(e^{\gamma_i f_t} | \Theta) E(e^{\gamma_j f_t} | \Theta) \right]. \end{aligned}$$

où $C = e^{\alpha_i + \alpha_j + \beta_i x_{it} + \beta_j x_{jt}}$. Comme on suppose que la variable f_t suit une normale standard on a les choses suivantes en considérant comme fixe les paramètres γ de Θ :

- $\gamma_i f_t \sim N(0; \gamma_i^2)$.
- $\gamma_j f_t \sim N(0; \gamma_j^2)$.
- $(\gamma_i + \gamma_j) f_t \sim N(0; (\gamma_i + \gamma_j)^2)$.

L'exponentielle d'une variable de loi normale $(\mu; \sigma^2)$ est une log-normale dont l'espérance est $e^{\mu + \frac{\sigma^2}{2}}$. On en déduit alors :

$$\begin{aligned} Cov(y_{it}^P, y_{it}^P | \Theta) &= C \left[e^{\frac{(\gamma_i + \gamma_j)^2}{2}} - e^{\frac{\gamma_i^2}{2}} e^{\frac{\gamma_j^2}{2}} \right] \\ &= C e^{\gamma_i \gamma_j} \\ &= C \exp(\gamma_i \gamma_j). \end{aligned}$$

Dans la suite de cette section, nous approximations grâce à la méthode delta une valeur approchée de $E(\lambda_{it})$ dans notre modèle linéaire généralisé. Le but est de montrer que nos λ_{it} générés seront assez grands afin que pour un même individu, i fixé, les valeurs y_{it}^P ne soient pas toutes nulles (probabilité proche de 0). Avec les paramètres des différentes distributions on recherche l'espérance et la variance de λ_{it} . Ici tout les paramètres du modèle sont considérés comme aléatoire. En effet, nous cherchons à vérifier que le processus de simulation génère des paramètres assez grands en moyenne. On rappelle que le modèle est :

$$\lambda_{it} = \exp(\alpha_i + \beta_i x_{it} + \gamma_i f_t)$$

et nous notons $\Lambda_{it} = \alpha_i + \beta_i x_{it} + \gamma_i f_t$.

Notons les choses suivantes $T = (\alpha_i, \beta_i, x_{it}, \gamma_i, f_t) = (T_1, T_2, T_3, T_4, T_5)$ le vecteur des 5 variables aléatoires et $\theta = (E(\alpha_i), E(\beta_i), E(x_{it}), E(\gamma_i), E(f_t)) = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$.

Puis on notera par g la fonction exponentielle : $g(T) = \exp(\alpha_i + \beta_i x_{it} + \gamma_i f_t) = \exp(\Lambda_{it})$. Enfin on notera les dérivées partielles de g de la manière suivante :

$$g'_k = \frac{\partial g}{\partial T_k},$$

où les T_i représentent les différentes variables aléatoires utilisées. On rappelle que le développement de Taylor s'applique sur des fonctions continues et dérivables, ce qui est le cas ici avec la fonction exponentielle. Le développement de Taylor d'ordre un de la fonction g autour de θ sera alors :

$$g(T) \approx g(\theta) + \sum_{i=k}^5 g'_k(\theta)(T_k - \theta_k). \quad (3.3.1)$$

En prenant l'espérance on a alors :

$$E(g(T)) \approx g(\theta) + \sum_{i=k}^5 g'_k(\theta)E(T_k - \theta_k)$$

et donc

$$E(g(T)) \approx g(\theta).$$

Ensuite on peut voir que

$$\begin{aligned} V(g(T)) &= E\left((g(T) - E(g(T)))^2\right) \\ &\approx E\left((g(T) - g(\theta))^2\right) \\ &\approx E\left(\left(g(\theta) + \sum_{k=1}^5 g'_k(\theta)(T_k - \theta_k) - g(\theta)\right)^2\right) \\ &\approx \sum_{k=1}^5 (g'_k)^2(\theta)V(T_k) + \sum_{k=1}^5 \sum_{l=1; k \neq l}^5 g'_k(\theta)g'_l(\theta)E((T_k - \theta_k)(T_l - \theta_l)) \\ &\approx \sum_{k=1}^5 (g'_k)^2(\theta)V(T_k). \end{aligned}$$

En effet, les termes $E((T_k - \theta_k)(T_l - \theta_l))$ sont nuls ici puisqu'ils représentent la covariance entre les différentes variables qui sont toutes indépendantes. Maintenant, on réalise les différentes dérivées partielles $g'_1(T) = \frac{\partial g}{\partial \alpha_i}(T) = \exp(\Lambda_{it})$, $g'_2(T) = \frac{\partial g}{\partial \beta_i}(T) = x_{it} \exp(\Lambda_{it})$, $g'_3(T) = \frac{\partial g}{\partial x_{it}}(T) = \beta_i \exp(\Lambda_{it})$, $g'_4(T) = \frac{\partial g}{\partial \gamma_i}(T) = f_t \exp(\Lambda_{it})$, $g'_5(T) = \frac{\partial g}{\partial f_t}(T) = \gamma_i \exp(\Lambda_{it})$.

On rappelle également les choses suivantes : $E(\alpha_i) = E(\beta_i) = \frac{1}{2}$, $V(\alpha_i) = V(\beta_i) = \frac{1}{12}$, $E(x_{it}) = 2$, $V(x_{it}) = 1$, $E(f_t) = 0$, $V(f_t) = 1$, $E(\gamma_i) = \frac{a+b}{2}$, $V(\gamma_i) = \frac{(b-a)^2}{12}$. On aura finalement :

$$E(\exp(\Lambda_{it})) \approx \exp\left(\frac{1}{2} + \frac{1}{2} \times 2\right) \approx 4,5$$

et

$$\begin{aligned} V(\exp(\Lambda_{it})) &\approx \exp\left(2 \times \left(\frac{1}{2} + \frac{1}{2} \times 2\right)\right) \\ &\times \left(\frac{1}{12} + 2 \times \frac{1}{12} + \frac{1}{2} \times 1 + \frac{a+b}{2}\right) \\ &\approx 20,01 \times \left(0,75 + \frac{(a+b)}{2}\right). \end{aligned}$$

Pour faire suite à nos calculs sur les covariances théoriques de la section précédente dans les modèles linéaires, on va utiliser ici une méthode similaire à celle utilisée pour $E(\lambda_{it})$ et $V(\lambda_{it})$ pour approximer les covariances des différents y_{it}^P lorsque l'on considère tout les paramètres comme aléatoires. En notant S de la même manière que T , $S = (\alpha_j, \beta_j, x_{js}, \gamma_j, f_s) = (S_1, S_2, S_3, S_4, S_5)$ le vecteur des 5 variables aléatoires et

$$\delta = (E(\alpha_j), E(\beta_j), E(x_{js}), E(\gamma_j), E(f_s)) = (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5).$$

Nous avons en utilisant la méthode delta :

$$\begin{aligned}
Cov(g(T), g(S)) &= E((g(T) - E(g(T)))(g(S) - E(g(S)))) \\
&\approx E((g(T) - g(\theta))(g(S) - g(\delta))) \\
&\approx E\left(\left(g(\theta) + \sum_{k=1}^5 g'_k(\theta)(T_k - \theta_k) - g(\theta)\right)\right. \\
&\quad \left.\times \left(g(\delta) + \sum_{l=1}^5 g'_l(\delta)(S_l - \delta_l) - g(\delta)\right)\right) \\
&\approx E\left(\left(\sum_{k=1}^5 g'_k(\theta)(T_k - \theta_k)\right)\left(\sum_{l=1}^5 g'_l(\delta)(S_l - \delta_l)\right)\right) \\
&\approx \sum_{k=1}^5 g'_k(\theta) \sum_{l=1}^5 g'_l(\delta) E((T_k - \theta_k)(S_l - \delta_l)) \\
&\approx \sum_{k=1}^5 \sum_{l=1}^5 g'_k(\theta) g'_l(\delta) Cov(T_k, S_l).
\end{aligned}$$

Rappelons que $\Lambda_{it} = \alpha_i + \beta_i x_{it} + \gamma_i f_t$ et $\Lambda_{js} = \alpha_j + \beta_j x_{js} + \gamma_j f_s$. En prenant, $i \neq j$ et $t = s$ on peut voir que le seul terme de covariance $Cov(T_k, S_l)$ non nul est $Cov(T_5, S_5) = Cov(f_t, f_t) = V(f_t)$. Ainsi on peut conclure que :

$$\begin{aligned}
Cov(g(T), g(S)) &= Cov(\exp(\Lambda_{it}), \exp(\Lambda_{jt})) \approx g'_5(\theta) g'_5(\delta) V(f_t) \\
&\approx E(\gamma_i)^2 \exp\left(\frac{1}{2} + \frac{1}{2} \times 2\right) \\
&\approx \left(\frac{a+b}{2}\right)^2 \exp(1,5) \\
&\approx 4,5 \left(\frac{a+b}{2}\right)^2.
\end{aligned}$$

On peut donc en déduire une approximation pour $Cov(\lambda_{it}, \lambda_{jt})$. On se rappelle ensuite que :

$$Cov(y_{it}^P, y_{jt}^P) = Cov(\lambda_{it}, \lambda_{jt}).$$

On remarquera que le résultat dépend bien de γ_i . D'une manière similaire on peut rechercher $Cov(y_{it}^P, y_{is}^P)$ pour $t \neq s$. En faisant l'inventaire des différentes variables aléatoires impliquées, les termes de covariances $Cov(T_k, S_l)$ non nuls sont $V(T_1) = V(\alpha_i), V(T_2) = V(\beta_i)$ et $V(T_4) = V(\gamma_i)$. On aura donc :

$$\begin{aligned}
Cov(\exp(\Lambda_{it}), \exp(\Lambda_{is})) &\approx g'_1(\theta)g'_1(\delta)V(\alpha_i) + g'_2(\theta)g'_2(\delta)V(\beta_i) + g'_4(\theta)g'_4(\delta)V(\gamma_i) \\
&\approx \exp(1,5)V(\alpha_i) + E(x_{it})E(x_{is}) \exp(2 \times 1,5)V(\beta_i) \\
&\quad + E(f_t)E(f_s) \exp(2 \times 1,5)V(\beta_i) \\
&\approx \exp(1,5)\frac{1}{12} + 2 \times 2 \times \exp(2 \times 1,5)\frac{1}{12} \\
&\approx 7,125.
\end{aligned}$$

On remarquera que cette quantité ne dépend pas des paramètres de γ_i utilisés.

CONCLUSION

Nous avons tout au long de ce mémoire étudié les performances des tests de Pesaran (2004), Friedman (1937) et Frees (1995). Après les avoir présentés dans le premier chapitre nous avons comparé les niveaux et puissances sur deux modèles linéaires. Dans le cas du modèle linéaire de Pesaran nous avons vu que le test de Pesaran, utilisant la statistique du CD , est le plus performant lorsque les corrélations sont toujours du même signe (positif dans nos simulations). Dans ce même cas, le test de Friedman a des performances très proches mais propose une alternative non paramétrique. Le test de Frees, quant à lui, présente des problèmes lorsque T est trop petit par rapport à N dans ce modèle.

Nous avons par la suite comparé ces tests avec un modèle linéaire simplifié sans le terme d'autorégression du modèle de Pesaran. Dans ce cas, le test de Frees démontre moins de problèmes lorsque N est plus grand que T mais la puissance est très faible lorsque $T = 5$. Utiliser le vrai quantile de la loi Q ne présente qu'une légère amélioration (voir annexe C). Le test de Pesaran reste le plus performant lorsque les corrélations sont du même signe. Afin de vérifier la suggestion de Frees (2004), nous avons ensuite comparé dans le modèle simplifié les performances des tests lorsque les corrélations alternent de signe. Dans ce cas le test de Frees est clairement à son avantage ayant une puissance plus importante que les deux autres. Le test de Pesaran continue d'être plus puissant que le test de Friedman dans ce cas.

Enfin dans le dernier chapitre, en s'inspirant de Hsiao *et al.* (2007) nous avons appliqué les trois tests avec un modèle linéaire généralisé de Poisson. Nous avons pu voir que les tests présentent des performances très similaires au cas linéaire. En effet, le test de Pesaran présente de très bonnes performances lorsque les corrélations sont toutes positives et une nouvelle fois le test de Frees est le moins performant en particulier lorsque $T = 5$. En utilisant une variable normale centrée en 0 pour le terme γ_i nous pouvons faire alterner le signe de la corrélation et dans ce cas le test de Frees est de nouveau à son avantage. On notera que comme mentionné dans ce chapitre on ne peut, avec notre modèle, avoir à la fois notre variable réponse y_{it}^P qui suit une Poisson et une dépendance entre les variables. Seulement $y_{it}^P | \lambda_{it}$ est Poisson mais dans ce cas nous ne pouvons avoir une dépendance entre les y_{it}^P par les λ_{it} . Une alternative serait d'utiliser la méthode des copules pour pouvoir modéliser la dépendance (voir par exemple Genest et Favre 2007).

Bibliographie

- Anselin, L. et A. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics textbooks and monographs* **155**, 237–290.
- Baltagi, B. (2005). *Econometric Analysis of Panel Data* (3e éd.). Hoboken, NJ : Wiley.
- Baltagi, B., S. Song, B. Jung et W. Koh (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* **140**, 5–51.
- Breusch, T. et A. Pagan (1980). The lagrange multiplier test and its application to model specifications in econometrics. *Review of Economic Studies* **47**, 239–253.
- Brown, G. et A. Mood (1951). On median tests for linear hypotheses. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 159–166.
- Cliff, A. et J. Ord (1981). *Spatial Processes. Models and applications*. Londres : Pion.
- Dobson, C. (2002). *An introduction to generalized linear models* (2e éd.). Boca Raton, FL : Chapman et Hall.
- Driscoll, J. et A. Kraay (1998). Consistent covariance matrix estimation with spatially dependent data. *Review of Economics and Statistics* **80**, 549–560.
- Fox, J. et W. Sanford (2011). *An R Companion to Applied Regression*. Thousand Oaks, CA : Sage Publications.

- Frees, E. (1995). Assessing cross-sectional correlation in panel data. *Journal of Econometrics* **69**, 393–414.
- Frees, E. (2004). *Longitudinal and Panel Data : Analysis and Applications in the Social Sciences*. Cambridge : Cambridge University Press.
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association* **32**, 675–701.
- Genest, C. et A.-C. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering* **12**, 347–368.
- Harper, J. (2003). A simple proof of $1 + \frac{1}{2^2} + \frac{1}{3^2} + \dots = \frac{\pi^2}{6}$ and related identities. *The American Mathematical Monthly* **109**(6), 540–541.
- Hollander, M. et D. Wolfe (1973). *Nonparametric statistical methods* (3e éd.). Hoboken, NJ : Wiley.
- Hoyos, R. D. et V. Sarafidis (2006). Testing for cross-sectional dependence in panel-data models. *The Stata Journal* **6**, 482–496.
- Hsiao, C., M. Pesaran et A. Pick (2007). Diagnostic tests of cross section independence for nonlinear panel data models. *Cambridge Working Papers in Economics* (0716).
- McCullagh, P. et J. Nelder (1989). *Generalized Linear Models* (2e éd.). Boca Raton, FL : Chapman and Hall.
- Pesaran, M. (2004). General diagnostic tests for cross section dependence in panels. *Cambridge Working Papers in Economics* (0435).
- Sarafidis, V., T. Yamagata et D. Robertson (2009). A test of cross section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics* **148**(2), 149–161.
- Sen, P. (1967). Asymptotically efficient tests by the method of n rankings. *Journal of the Royal Statistical Society. Series B (Methodological)* **504**(2), 312–317.

Annexe A

TEST DES DIFFÉRENTS TYPES DE RÉSIDUS POUR MODÈLE LINÉAIRE GÉNÉRALISÉ

Cette annexe fait écho au chapitre trois sur l'utilisation des tests avec les résidus sur un modèle linéaire généralisé. Puisqu'il y a un choix à faire sur les résidus de tels modèles, nous exposons dans cette annexe les résultats obtenus avec les résidus de type réponse, de Pearson, et de la déviance. On rappelle tout d'abord la forme de ces résidus selon Dobson (2002) et Hsiao *et al.* (2007) en notant $\hat{y}_{it}^P = E(y_{it}^P | x_{it}, \hat{\theta}_i)$ les valeurs prédites (estimation de l'espérance conditionnelle de y_{it}^P). On notera que pour un modèle linéaire généralisé de Poisson $\hat{y}_{it}^P = V(y_{it}^P | x_{it}, \hat{\theta}_i) = \hat{\lambda}_{it}$ Les résidus de Pearson sont :

$$\frac{y_{it}^P - \hat{y}_{it}^P}{\sqrt{V(y_{it}^P | x_{it}, \hat{\theta}_i)}},$$

ceux de la déviance

$$\text{Signe}(y_{it}^P - \hat{y}_{it}^P) \sqrt{d_{it}},$$

où $d_{it} = \sqrt{2y_{it}^P \log\left(\frac{y_{it}^P}{\hat{y}_{it}^P}\right) - 2(y_{it}^P - \hat{y}_{it}^P)}$ est la contribution à la déviance de l'observation y_{it}^P . Si y_{it}^P est nul alors le \log dans d_{it} est évalué à 0. Enfin les résidus de type réponse sont simplement $y_{it}^P - \hat{y}_{it}^P$.

A-ii

Dans les tableaux suivants, on compare quelques résultats avec les trois formes de résidus que nous avons essayées. Les simulations sont faites avec 1000 répétitions et le modèle utilisé dans le chapitre trois. Pour γ_i suivant une uniforme, nous avons vu que le meilleur test est celui de Pesaran. Nous avons donc testé les différents types de résidus uniquement pour ce test.

TABLE A.1. Puissance, modèle linéaire généralisé, $\gamma_i \sim U(0,1;0,3)$. Résidus de type Réponse test du *CD*

N \ T	5	10	25	50	100	200	300
5	0,166	0,286	0,556	0,791	0,927	0,994	0,996
10	0,325	0,619	0,916	0,993	1,000	1,000	1,000
25	0,635	0,944	1,000	1,000	1,000	1,000	1,000
50	0,846	0,994	1,000	1,000	1,000	1,000	1,000
100	0,931	0,998	1,000	1,000	1,000	1,000	1,000
200	0,981	1,000	1,000	1,000	1,000	1,000	1,000
300	0,989	1,000	1,000	1,000	1,000	1,000	1,000

TABLE A.2. Puissance, modèle linéaire généralisé, $\gamma_i \sim U(0,1;0,3)$. Résidus de type Pearson test du *CD*

N \ T	5	10	25	50	100	200	300
5	0,158	0,282	0,576	0,802	0,931	0,997	0,998
10	0,328	0,612	0,920	0,992	1,000	1,000	1,000
25	0,632	0,946	1,000	1,000	1,000	1,000	1,000
50	0,847	0,994	1,000	1,000	1,000	1,000	1,000
100	0,929	0,998	1,000	1,000	1,000	1,000	1,000
200	0,982	1,000	1,000	1,000	1,000	1,000	1,000
300	0,988	1,000	1,000	1,000	1,000	1,000	1,000

TABLE A.3. Puissance, modèle linéaire généralisé, $\gamma_i \sim U(0,1;0,3)$. Résidus de type Déviance test du *CD*

N \ T	5	10	25	50	100	200	300
5	0,168	0,287	0,562	0,805	0,939	0,998	0,998
10	0,333	0,615	0,927	0,991	1,000	1,000	1,000
25	0,640	0,951	1,000	1,000	1,000	1,000	1,000
50	0,851	0,995	1,000	1,000	1,000	1,000	1,000
100	0,930	0,999	1,000	1,000	1,000	1,000	1,000
200	0,979	1,000	1,000	1,000	1,000	1,000	1,000
300	0,988	1,000	1,000	1,000	1,000	1,000	1,000

Comme on peut le voir dans les tableaux précédents, peu importe les résidus utilisés les résultats sont très similaires.

Annexe B

RÉSUMÉ DES MODÈLES UTILISÉS POUR LES SIMULATIONS

Dans tous les modèles nous utilisons :

- (1) Aucune dépendance : $\gamma_i = 0$,
- (2) Faible dépendance : $\gamma_i \sim \text{Siid } U(0,1;0,3)$,
- (3) Moyenne dépendance : $\gamma_i \sim \text{Siid } U(0,2;0,4)$,
- (4) Forte dépendance : $\gamma_i \sim \text{Siid } U(0,3;0,5)$.
- (5) Dépendance centrée en 0 : $\gamma_i \sim \text{Siid } N(0;1)$.

Le modèle linéaire dans Pesaran (2004) est le suivant :

$$y_{it}^A = \mu_i(1 - \beta_i) + \beta_i y_{i(t-1)} + u_{it}.$$

Afin de rendre compte de la dépendance, les erreurs sont simulées à l'aide du modèle suivant :

$$u_{it} = \gamma_i f_t + \epsilon_{it}, i = 1, 2, \dots, N; t = 1, 2, \dots, T,$$

où $\epsilon_{it} \sim \text{iid } N(0,1)$, $\beta_i \sim \text{Siid } U(0;1)$, $f_t \sim \text{iid } N(0,1)$, $\mu_i \sim \text{Siid } N(0,2)$.

Le modèle linéaire simplifié que l'on utilise est :

$$y_{it}^L = \alpha_i + \beta_i x_{it} + u_{it}.$$

B-ii

Les erreurs u_{it} sont simulées de la même façon que précédemment. On prend $\alpha_i \sim \text{Siid } U(0;1)$, $\beta_i \sim \text{Siid } U(0;1)$ et $x_{it} \sim \text{Siid } N(2,1)$.

Le modèle linéaire généralisé utilisé ici est :

$$\log(\lambda_{it}) = \alpha_i + \beta_i x_{it} + \gamma_i f_t$$

et $y_{it}^P | \lambda_{it}$ suit une loi de Poisson de paramètre λ_{it} , $x_{it} \sim \text{Siid } N(2;1)$, $\alpha_i \sim \text{Siid } U(0;1)$, $\beta_i \sim \text{Siid } U(0;1)$. Les termes γ_i et f_t sont comme dans les modèles linéaires.

Annexe C

TEST DE FREES POUR T=5

Dans nos simulations des chapitres deux et trois, nous avons utilisé l'approximation de la loi Q par une loi normale. Comme mentionné dans la section 1.3, cette approximation est médiocre pour $T = 5$. Nous présentons dans cette annexe des simulations refaites en comparant les résultats obtenus avec cette approximation normale et ceux obtenus en utilisant le vrai quantile 95% de la loi Q disponible dans Frees (2004). Ce quantile prend la valeur 0,683. Pour le modèle de Pesaran linéaire nous obtenons les niveaux suivants :

TABLE C.1. Niveaux, modèle linéaire de Pesaran $\gamma_i = 0$, comparaisons pour le test de Frees lorsque T=5

N	5	10	25	50	100
Avec le quantile Q	0,095	0,100	0,206	0,398	0,759
Avec l'approximation	0,083	0,090	0,189	0,369	0,745

Comme on peut le voir nous n'obtenons pas d'amélioration nette pour les niveaux ou les puissances. Les niveaux continuent d'être mauvais pour le modèle de Pesaran. Il est inutile de faire les puissances étant donné que les niveaux ne sont pas dans un intervalle acceptable. Pour le modèle linéaire simplifié nous obtenons les niveaux suivants :

TABLE C.2. Niveaux, modèle linéaire simplifié $\gamma_i = 0$, comparaisons pour le test de Frees lorsque T=5

N	5	10	25	50	100
Avec le quantile Q	0,089	0,060	0,054	0,042	0,050
Avec l'approximation	0,072	0,053	0,045	0,036	0,047

C-ii

et les puissances suivantes pour $\gamma_i \sim U(0,3;0,5)$:

TABLE C.3. Puissance, modèle linéaire simplifié $\gamma_i \sim U(0,3;0,5)$, comparaisons pour le test de Frees lorsque T=5

N	5	10	25	50	100
Avec le quantile Q	0,070	0,057	0,097	0,157	0,224
Avec l'approximation	0,061	0,051	0,086	0,148	0,210

Les niveaux étaient déjà bons avec l'approximation par la loi normale (sauf pour $N = 5$). Ils le sont encore en prenant le bon quantile. On observe seulement une petite amélioration de la puissance en prenant ce quantile plutôt que l'approximation.

En fait, il est assez simple de voir que le quantile de la loi normale approximant Q pour $T = 5$ est $1,96 \times \sqrt{V(Q)} = 0,717$ et donc puisque ce quantile est supérieur à celui de la vraie distribution de Q (0,683) on rejettera moins souvent (niveaux plus bas, puissances plus faible). L'approximation est donc conservatrice dans ce cas.