UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LA RÉGRESSION EXPECTILE

POUR L'ANALYSE DES DONNÉES LONGITUDINALES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN MATHÉMATIQUES

PAR

AMADOU DIOGO BARRY

MARS 2019

REMERCIEMENTS

J'ai lu pour la première fois sur la régression quantile quand je travaillais à l'Institut national de la santé publique du Québec (INSPQ). À cette époque, j'avais terminé ma maîtrise et je commençais ma carrière en statistiques. Je me suis dit WOW, un boxplot multivarié et une foule de questions me sont venues en tête. Pourquoi n'est-ce pas enseigné dans les cours de régression, pourquoi cette méthode n'est-elle pas intégrée au logiciel SAS (je connaissais moins R, qui ne me permettait pas de traiter des fichiers de données volumineux comme les fichiers administratifs de naissances et de décès du Québec). Naturellement, j'ai voulu en savoir un peu plus sur la régression quantile. En poussant mes recherches, je suis tombé sur le blogue du professeur Arthur Charpentier qui avait publié un excellent billet sur le sujet. Pour le coup, tous ses codes étaient en R, alors j'ai téléchargé le logiciel R. J'ai aimé son approche, alors je suis devenu un lecteur assidu de son blogue. J'estime que Dr Charpentier a une approche élégante et qu'il sait vulgariser efficacement auprès du grand public des concepts statistiques et actuariels poussés, sans occulter la théorie mathématique et statistique qui formalise ces concepts.

Je me rappelle encore un souper avec ma conjointe pendant lequel je lui parlais d'un billet que j'avais lu pendant la journée. Ma conjointe m'a fait remarquer que je m'enthousiasmais plus pour le blogue que pour mon travail. Elle m'a dit : "tu devrais contacter monsieur Charpentier pour qu'il soit ton directeur de thèse". C'est ainsi que le 23 février 2012 j'ai sollicité Dr Charpentier pour qu'il me dirige au doctorat. Après

quelques échanges, il a accepté et nous avons amorcé notre collaboration. Certes, je le remercie profondément pour sa disponibilité, ses conseils et ses remarques. Mais, avant toute chose, je le remercie de m'avoir accordé sa confiance et de m'avoir permis de me lancer dans cette aventure doctorale, un rêve que je chéris depuis des années.

Après quelques années, Dr Charpentier est parti en France. J'ai donc dû trouver un nouveau directeur à l'UQAM. La première personne à qui j'ai pensé fut le professeur Karim Oualkacha (qui porte bien son nom, en passant). J'avais suivi un cours sur la génétique avec lui, et j'avais trouvé qu'il donnait beaucoup de travail pour un cours valant 1 seul crédit ! Or, le cours était très intéressant et l'approche du professeur d'autant plus instructive. Dr Oualkacha a accepté de me co-diriger et j'ai beaucoup appris à ses côtés. Je ne pourrai jamais assez le remercier pour sa direction. À chacune de nos rencontres, je ressortais la tête remplie de choses nouvelles, d'astuces, de références, etc. En bref, Dr Oualkacha est un excellent pédagogue, soucieux que je prenne le temps d'expliquer tout dans les moindres détails. D'ailleurs, pour l'anecdote, je me souviens d'un entretien avec lui, de près de 4 heures, pendant lequel j'ai écrit des formules au tableau. Je le remercie pour sa disponibilité et pour son implication exemplaires.

Je tiens à remercier également tous les membres du jury qui m'ont fait l'honneur de lire ma thèse et de l'évaluer. Je vous remercie pour votre temps précieux et vos commentaires pertinents.

Je tiens à remercier les professeurs du département de mathématiques à l'UQAM que j'ai connus en classe ou simplement dans les corridors.

Merci au personnel administratif du département, en particulier à Isabelle Couture pour son assistance sur les questions administratives et à Gisèle Legault pour son soutien technique, notamment pour Latex.

Cette aventure mémorable a eu lieu, en partie, grâce à ma conjointe. Elle a toujours été là pour me soutenir pendant les périodes difficiles. Je la remercie pour tous les soirs et les week-ends pendant lesquels j'ai pu travailler sur ma thèse, alors qu'elle s'occupait de nos enfants. En vérité, elle m'a soutenu même avant la naissance de nos enfants. Elle a toujours été là avec son enthousiasme et son énergie positive. Heureuse combinaison professionnelle, elle a pu m'aider avec les tâches de révision en anglais et en français.

Naturellement, je ne peux oublier les remerciements essentiels qui reviennent à mes parents, qui m'ont toujours soutenu et pour qui cette réalisation à une signification sans égal. Sans eux, cet accomplissement n'aurait jamais eu ni de début ni de fin. À vous je dis : On jaramaa, babaa et nenee, yo allah yobohon, ukoraon chelallal, achidaé ha boya djouta.

Finalement, je remercie le Fonds de recherche Société et culture, ainsi que le Fonds de recherche Nature et technologies pour leur soutien financier indispensable. Je remercie également le Centre interuniversitaire québecois de statistiques sociales (CIQSS) pour l'octroi d'une bourse d'accueil.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

NOTATION

Les vecteurs sont écrits en gras minuscule, $\boldsymbol{a} \in \mathbb{R}^n$. Les matrices sont écrites en gras majuscule $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. Les quantités estimées à partir des données sont écrites avec des chapeaux, $\widehat{\boldsymbol{A}}$. Le transposé et l'inverse de la matrice $\boldsymbol{A}$ sont notés respectivement par : $\boldsymbol{A}^\intercal$, $\boldsymbol{A}^{-1}$. Le vecteur $\mathbb{1}_q$ est le vecteur unitaire, le vecteur $\mathbf{0}$ est le vecteur nul et la matrice identité de taille $q \times q$ est notée par : $\mathbb{I}_q$.

RÉSUMÉ

La régression, au sens large, est l'une des méthodes d'inférence les plus utilisées en modélisation. La régression modélise la relation entre les régresseurs et la variable réponse. Cette modélisation se résume par l'estimation de l'influence des régresseurs sur la moyenne conditionnelle de la variable réponse. Alors que l'inférence sur la moyenne conditionnelle est généralement acceptable, il arrive que l'intérêt porte sur l'estimation des queues de la distribution de la variable réponse conditionnellement aux régresseurs. Dans ce contexte, la régression classique est inefficace et il faut aller au-delà de l'estimation de la moyenne conditionnelle. La littérature moderne offre des approches pour répondre à ce genre de problématique, notamment avec la régression asymétrique des moindres carrés pondérés.

La régression asymétrique des moindres carrés pondérés ou régression expectile (RE) a récemment gagné en popularité, en partie grâce à ses propriétés statistiques et computationnelles attrayantes. La RE estime les fonctions expectiles/percentiles de la distribution de la variable réponse en fonction des régresseurs et de leur coefficient. Par conséquent, la RE permet d'examiner et d'analyser l'influence des régresseurs sur la distribution conditionnelle de la variable réponse, révélant ainsi une variété de formes d'hétérogénéité. De plus, la RE est très simple à mettre en oeuvre comparativement à son analogue, la régression quantile (RQ).

Dans la présente thèse, nous introduisons la RE à l'analyse des données longitudinales. Nous étudions l'association de la RE au modèle GEE et au modèle linéaire avec effets-fixes (EF). Le modèle GEE et le modèle EF sont des modèles très réputés et communément utilisés en biostatistique et en économétrie. Les données longitudinales sont de loin les données observationnelles les plus appréciées. Les données longitudinales prennent en compte la dynamique, le développement et le changement de la population à l'étude et offrent une meilleure inférence des paramètres du modèle. Ensuite, nous présentons le plan de la thèse.

Dans le chapitre préliminaire, Chapitre I, nous introduisons les statistiques asymétriques (quantile et expectile) et quelques-unes de leurs propriétés. Nous discutons leurs similarités et complémentarités. Par la suite, nous introduisons les modèles de la régression quantile (RQ) et de la régression expectile (RE) associés au modèle linéaire simple. Après l'introduction des modèles RQ et RE, nous présentons les propriétés asymptotiques de leur estimateur. Nous terminons le chapitre par la présentation succincte du modèle GEE, du modèle EF et du modèle linéaire avec effets-aléatoires (EA), ainsi que les propriétés asymptotiques de leur estimateur.

Dans le second chapitre (Chapitre II), nous introduisons une nouvelle classe d'estimateurs qui découle de l'association de la régression des moindres carrés asymétriques pondérés et des équations d'estimation généralisées (GEE). Cette nouvelle classe estime l'expectile de la variable réponse en fonction des régresseurs et inclut une structure de corrélation hypothétique dans les équations d'estimation pour modéliser la dépendance des données. De plus, les structures de corrélation couramment utilisées avec le modèle GEE se généralisent et s'appliquent naturellement dans les équations d'estimation de cette nouvelle classe d'estimateurs. Cette dernière permettra au modèle GEE de capturer l'hétérogénéité des effets des régresseurs et de tenir compte de l'hétérogénéité non observée.

Nous avons montré les propriétés asymptotiques de ces nouveaux estimateurs et avons proposé un estimateur robuste de leur matrice de variance-covariance. Les résultats des simulations exhaustives ont démontré leurs qualités favorables dans différents scénarios et leurs avantages par rapport à d'autres méthodes similaires. Finalement, nous avons étudié l'effet d'un nouveau traitement sur la douleur du travail pendant l'accouchement pour illustrer la méthode.

Le troisième chapitre (Chapitre III) introduit le modèle de la régression expectile avec effets-fixes (ERFE). Le modèle ERFE hérite de propriétés attrayantes pour l'analyse des données longitudinales. D'abord comme extension du modèle EF, le modèle ERFE, dans sa spécification, tient compte de la corrélation entre les régresseurs du modèle et les caractéristiques individuelles non-observées, comme les facteurs génétiques et environnementaux. Ensuite, grâce à l'approche de la régression des moindres carrés asymétriques pondérés, le modèle ERFE permet l'estimation et l'analyse de l'influence des régresseurs sur la localisation, l'échelle et la forme de la distribution conditionnelle de la variable réponse.

Cela dit, le modèle ERFE pose aussi le problème lié au modèle EF désigné par les termes «incidental parameter problem».
Nous montrons que l'estimateur ERFE est un «iterative within-transformation estimator». Autrement dit, l'estimateur ERFE peut être dérivé en utilisant de manière itérative la stratégie de la «within-transformation» proposée dans le cadre du modèle EF pour résoudre le problème et éliminer le paramètre individuel. Nous établissons les propriétés asymptotiques de l'estimateur ERFE et suggérons un estimateur convergent et hétéroscédastique pour sa matrice de variance-covariance.

Nous avons évalué les performances de l'estimateur ERFE à travers une simulation exhaustive et l'avons comparé au modèle de la régression quantile avec effets-fixes (QRFE). Les résultats sont mitigés, le modèle ERFE est compétitif et plus performant dans certains scénarios. Nous l'avons employé pour étudier le rendement scolaire sur le salaire avec les données réelles sur l'étude de la dynamique des revenus (PSID).

Le dernier chapitre (Chapitre IV) porte sur une approche originale pour résoudre le «incidental parameter problem» dans le modèle ERFE. Cette approche, que nous désignons par PERFE, consiste à appliquer une pénalité au paramètre individuel. En plus de conserver les propriétés attrayantes du modèle ERFE, le modèle PERFE permet l'estimation des régresseurs invariants dans le temps. Nous avons appliqué la pénalité l_1 afin de régulariser le paramètre individuel autour de la valeur zéro. Le degré de régularisation est contrôlé par le paramètre de régularisation et sa valeur optimale est choisie en s'appuyant sur le critère d'information bayésien (BIC). Nous appliquons également une astuce pour déterminer le chemin de la solution du paramètre de régularisation et réduire le temps de calcul.

Les résultats de la simulation montrent que l'estimateur PERFE est plus performant que le modèle ERFE et le modèle QRFE avec pénalité (PQRFE). Nous appliquons le modèle PERFE aux données PSID pour étudier l'hétérogénéité du rendement scolaire.

MOTS-CLÉS : Expectile regression, Quantile regression, Longitudinal data, Panel data, GEE, Working correlation, Fixed-effets, Within-transformation, Penalty method, shrinkage.

INTRODUCTION

La régression statistique est l'une des méthodes de modélisation la plus ancienne disponible dans la boîte à outils de l'analyste de données. On peut retracer ses origines au milieu du XIX siècle dans les travaux de Boskovic qui, en 1755-1757, minimise la somme des valeurs absolues des erreurs dans le but de mesurer la longueur de cinq méridiens terrestres. Quelques années plus tard, en 1789, Laplace lui emboîte le pas et utilise la même approche dans son célèbre traité (Lovering, 1888). Une deuxième méthode de régression communément désignée *méthode des moindres carrés ordinaires (MCO)* fut quant à elle introduite au début du XIX siècle. La méthode MCO minimise la somme des carrés des résidus. À l'époque, la méthode MCO était la méthode statistique la plus accessible compte tenu des limitations computationnelles. Cette méthode est attribuée à Legendre et à Gauss qui l'ont développée indépendamment et l'ont utilisée dans le domaine de la physique et de l'astronomie. C'est cette dernière approche qui constitue la régression statistique, telle qu'on la connaît aujourd'hui. L'expression «régression linéaire» et le développement de sa fondation mathématique ont fait leur apparition plus tard, dans le domaine de la biostatistique. En effet, l'expression «régression linéaire» apparaît pour la première fois en 1886, dans l'illustre article de Galton (1886) sur l'étude de la relation entre la taille des fils et celle des pères. Pearson, le collègue de Galton, a rigoureusement élaboré la théorie mathématique à l'origine de la régression multiple, utilisée aujourd'hui.

Présentement, la régression classique et ses différentes extensions sont utilisées dans plusieurs domaines des sciences appliquées, notamment en écologie, génétique, économie, sociologie, assurances et en médecine. La régression classique modélise la relation entre les régresseurs et la variable réponse. Cette relation, entre la variable réponse et les variables explicatives, est résumée par l'estimation de l'effet de ces dernières sur la moyenne de la distribution de la variable réponse. La moyenne est une mesure de tendance centrale avec de belles propriétés statistiques. Par exemple, la moyenne est une statistique exhaustive pour la famille exponentielle et elle est facilement calculable. Souvent, l'influence des régresseurs sur la variable réponse change selon qu'on se trouve à gauche ou à droite de la queue de la distribution de cette dernière. Dans de telles situations, l'effet dans la moyenne uniquement ne reflète pas cette hétérogénéité de l'effet des régresseurs. Considérons, par exemple, les facteurs de risque comme le statut tabagique de la mère pendant la grossesse sur le poids du bébé à la naissance. Il y a de fortes chances que l'impact et la magnitude de l'effet du tabagisme pendant la grossesse diffèrent pour un bébé en insuffisance de poids versus un bébé en embonpoint. Ainsi, la régression classique est impuissante ou insuffisante pour capturer l'effet des régresseurs sur les queues de la distribution de la variable réponse. À moins que l'effet des régresseurs soit uniforme sur toute la distribution de la variable réponse.

C'est dans ce contexte que d'autres méthodes de régression, capables d'estimer plus que la moyenne, ont vu le jour. Parmi ces méthodes, il y a la régression quantile (RQ) et la régression expectile (RE). Avec le progrès informatique et l'augmentation de la puissance de calcul, ces deux méthodes ont élargi le champ de la modélisation statistique au-delà des frontières de la régression classique de la moyenne. La RQ et la RE permettent d'estimer l'effet des régresseurs à différents points ou percentiles

de la fonction de répartition de la variable réponse. Plus précisément, la RQ estime les quantiles tandis que la RE estime les expectiles de la fonction de répartition de la variable réponse en fonction des variables explicatives et des paramètres du modèle. Les deux méthodes prennent en considération l'hétérogénéité non-observée et l'hétérogénéité des effets des variables explicatives. De plus, les deux méthodes n'énoncent aucune hypothèse sur la forme de la distribution de la variable réponse. La RQ et la RE génèrent de nouvelles classes d'estimateurs qui offrent un portrait global et détaillé de la variabilité de la variable réponse en fonction des variables explicatives du modèle. Le rôle de ces deux classes d'estimateurs, quantiles et expectiles, est très similaire dans la modélisation. En somme, l'estimation de l'une ou l'autre de ces classes suffit pour dresser un portrait global. Toutefois, elles se distinguent par leurs avantages et leurs inconvénients.

Dans la présente thèse, nous avons étudié particulièrement la RE. En raison de la similarité et de la complémentarité des deux méthodes, nous présentons la revue de littérature des deux méthodes. Nous mettons l'accent sur leur application aux données longitudinales.

Le quantile de niveau α est une statistique d'ordre communément utilisé en analyse descriptive. La médiane, les quartiles et les quintiles font partie des quantiles les plus familiers et les plus courants. Dans une analyse descriptive, ils offrent une image complète, sous forme de box-plot par exemple, de la distribution de la variable. Grâce à la RQ, il est possible de dresser un portrait similaire de la relation entre plusieurs variables en modélisation. Par exemple, nous pouvons théoriquement avoir deux box-plots de l'effet du tabagisme sur le poids du bébé, un pour les fumeurs et un autre pour les non-fumeurs.

L'introduction de la RQ remonte à 1978 avec l'article fondateur de Koenker et Bassett (1978). Dans cet article, les auteurs proposent une nouvelle classe d'estimateurs basée sur la RQ. Ils présentent les propriétés d'équivariance de cette classe et établissent leurs propriétés asymptotiques. L'originalité de l'article réside dans l'introduction des quantiles comme solution de minimisation d'une fonction de risque fondée sur la fonction de perte l_1. Comme cette fonction de risque n'est pas partout différentiable, les méthodes d'estimation traditionnelles, par exemple l'algorithme de Gauss-Newton, ne s'appliquent plus pour générer les estimateurs de la RQ. Pour contourner ce problème, Koenker et Bassett ont reformulé le problème d'optimisation dans le cadre de la programmation linéaire, une autre innovation de leur article, (Koenker et Bassett, 1978).

Les solutions obtenues par cette méthode ne sont pas analytiques, mais il est possible de les approcher numériquement avec différents algorithmes. L'algorithme du simplexe est le premier algorithme proposé pour trouver les estimateurs de la RQ. Une version plus efficace et plus attractive de cet algorithme fut proposée par Barrodale (1973). Un autre algorithme réputé plus efficace et plus attrayant en temps de calcul, lorsque la taille de l'échantillon augmente, est l'algorithme du point intérieur (Koenker et Vasco, 1987). Aujourd'hui avec le progrès technologique, la recherche de nouveaux algorithmes et l'amélioration de leur efficacité est devenue un champ de recherche très actif. En particulier, pour l'application de la RQ aux données de grandes dimensions. Pour obtenir plus de détails à ce sujet, veuillez consulter l'article de Mkhadri *et al.* (2017).

Pour favoriser l'accessibilité à cette nouvelle méthode et développer des algorithmes performants, la question de l'inférence statistique s'est posée. Koenker et Basset sont précurseurs dans ce domaine. Ils ont proposé 3 tests pour les estimateurs l_1, qui correspondent au test de Wald, au test des rapports de vraisemblance et au test du multiplicateur de Lagrange, (Koenker et Bassett, 1982). Par la suite, Koenker et Machado (1999) ont développé un indicateur de la qualité d'ajustement de la RQ linéaire similaire au coefficient de détermination utilisé pour l'estimateur de la régression des moindres carrés. Pour compléter l'inférence des estimateurs de la RQ, il ne restait plus que la génération d'intervalle de confiance. Sur ce plan, la tâche s'annonçait plus ardue. En effet, l'expression de la matrice de variance-covariance de l'estimateur de la RQ est fonction de la densité de la variable d'erreur et l'estimation de cette dernière est coûteuse en temps et en puissance de calcul. Koenker et Machado ont réalisé une revue de littérature sur le sujet (Koenker et Machado, 1999). Dans cet article, ils ont recensé les différentes approches d'estimation de la densité de l'erreur et plusieurs méthodes pour estimer directement l'intervalle de confiance. Dans ce même article, ils ont présenté d'autres alternatives pour générer les intervalles de confiance des estimateurs de la RQ, comme la méthode de l'inversion du test de rang ou la méthode du rééchantillonnage (Bootstrap). Cette dernière méthode existe selon différentes versions cataloguées dans l'article de Koenker et Machado (1999). En terme de rééchantillonnage, Kocherginsky *et al.* (2005) ont proposé un algorithme plus performent en temps pour générer les intervalles de confiance des estimateurs de la RQ à partir d'une chaîne de Markov marginale modifiée. Ce développement théorique a favorisé l'accessibilité et l'utilisation de la RQ. Aujourd'hui, son champ d'application couvre pratiquement tous les domaines des sciences appliquées. Pour une bonne revue de littérature sur le sujet, on peut consulter l'excellent livre de Koenker (2005) où la récente revue de littérature (Koenker *et al.*, 2018).

Finalement, la recherche sur la RQ s'est concentrée à son adaptation ou extension à différents modèles préalablement élaborés pour modéliser les effets des variables explicatives sur la moyenne de la variable réponse. Powell (1986) a généralisé la RQ aux données censurées, Machado et Silva (2005) l'ont adaptée aux données de comptage. Koenker et Bilias (2002), Fitzenberger et Wilke (2006) l'ont appliquée aux modèles de durées. Plus récemment, la RQ a été adaptée aux données longitudinales (Koenker, 2004; Canay, 2011; Harding et Lamarche, 2009; Lamarche, 2010; Galvao, 2011; Galvao et Montes-Rojas, 2010). L'article de Farcomeni et Marino (2015) et le chapitre 19 du livre «Handbook of quantile regerssion» (Koenker *et al.*, 2018, chap. 19) présentent une excellente revue de littérature de l'adaptation de la RQ aux différents modèles d'analyse des données longitudinales.

Contrairement au quantile, l'expectile n'est pas une statistique courante, hormis la moyenne qui est un expectile de niveau $\tau = 0.5$. L'expectile est une moyenne pondérée avec la seule particularité que son poids associé est une variable aléatoire fonction des données. Comme les quantiles, les expectiles suffisent pour décrire la distribution d'une variable aléatoire, notamment avec la moyenne et quelques expectiles supérieurs et inférieurs à la moyenne. Il est aussi possible de représenter la distribution d'une variable aléatoire avec un box-plot produit à partir des expectiles. Nous retrouvons dans la littérature deux écoles (Kneib, 2013b; Eilers, 2013; Koenker, 2013) qui opposent quantile et expectile et qui exhaussent les qualités de l'un par rapport à l'autre. À cet égard, nous les considérons complémentaires avec chacun ses avantages et ses inconvénients. Les expectiles généralisent la moyenne et sont plus faciles à calculer, tandis que les quantiles généralisent la médiane et sont plus robustes. Par ailleurs, mentionnons que le quantile est une statistique d'ordre

alors que l'expectile est une moyenne pondérée. En d'autres termes, les quantiles se concentrent sur l'ordre des observations tandis que les expectiles ciblent leurs valeurs (Farooq et Steinwart, 2017). Dans notre thèse, nous nous intéressons au modèle RE qui est une généralisation naturelle de la régression OLS, RE du niveau $\tau = 0.5$.

L'introduction du modèle de la régression expectile (RE) dans la littérature remonte à la même période que celle de la RQ (Koenker et Bassett, 1978). L'estimateur du modèle de la RE fut introduit en 1978 par Aigner *et al.* (1976) sous le nom de l'estimateur des moindres carrés asymétriques pondérés. Cependant, c'est dans l'article de Newey et Powell (1987) que le nom expectile fut évoqué pour la première fois. Outre l'introduction du nom, on retrouve dans cet article (Newey et Powell, 1987) une étude détaillée des propriétés de la fonction expectile. Notamment, le fait que la fonction expectile est invariante par translation affine. Cet article présente également la méthode de la RE pour un modèle linéaire simple et montre les propriétés asymptotiques de cette nouvelle classe d'estimateurs. Quelques années plus tard, Efron publie un article (Efron, 1991) illustrant les bienfaits de l'utilisation de cette nouvelle classe d'estimateurs tout en vantant ses propriétés analogues à celles des quantiles. Par la suite, peu ou pas d'études sur le modèle de la RE ont été publiées dans la littérature, probablement, parce que les efforts étaient concentrés sur le développement du modèle de la RQ. Comme le mentionne Eilers (2013), après l'article d'Efron (1991), le modèle de la RE est resté dans l'ombre pendant des décennies. Aujourd'hui, le sujet refait surface et intéresse les chercheurs comme le démontre le grand nombre de publications sur le sujet. Les premières contributions ont porté sur l'application de la méthode RE au modèle de spline et de lissage (Schnabel et Eilers, 2009; Rossi et Harvey, 2009; Sobotka et Kneib, 2012; Sobotka *et al.*, 2013). D'autres travaux ont été publiés, notamment sur la confrontation des deux statistiques. L'objectif visait à

montrer comment obtenir des quantiles à partir d'une grille fine d'expectiles. Finalement, des articles sont parus sur le problème du croisement des courbes et sur la promotion de l'application du modèle RE (Kneib, 2013a; Schnabel et Eilers, 2013; Waltrup *et al.*, 2015a).

Aujourd'hui, la RE est combiné à de nombreux types de modèles : bayésiens (Majumdar et Paul, 2016; Waldmann *et al.*, 2016; Xing et Qian, 2017), non-paramétriques (Righi *et al.*, 2014; Yang et Zou, 2015), non-linéaires (Kim et Lee, 2016), réseau de neurones (Xu *et al.*, 2016; Jiang *et al.*, 2017), machine à vecteurs de support (Farooq et Steinwart, 2017), lissage et spline avec pénalité (Schulze et Kauermann, 2015).

Notre thèse est présentée dans l'ordre suivant : l'amorce est constituée d'un chapitre préliminaire. Dans ce chapitre, nous présentons formellement les quantiles et les expectiles, puis la RQ et la RE. Ensuite, nous introduisons les données longitudinales ainsi que quelques méthodes développées particulièrement pour leur analyse. Les chapitres subséquents sont formés de trois articles différents dont la somme constitue le corps de la présente thèse. Finalement, la conclusion résume notre travail et apporte quelques pistes de réflexion sur de nouvelles avenues de recherche.

CHAPITRE I

CHAPITRE PRÉLIMINAIRE

1.1 Quantiles et Expectiles

Nous allons introduire le quantile et l'expectile d'une variable aléatoire Y dans le cas univarié. Nous présenterons les caractéristiques de chacun et nous examinerons leur ressemblance.

1.1.1 Quantiles

Le quantile de niveau $\theta \in [0,1]$ de la variable aléatoire Y est défini par

$$q_\theta(Y) = F^{-1}(\theta) = \inf\{y; F(y) \geq \theta\},$$

où F est la fonction de répartition de la variable aléatoire Y. Considérée comme fonction de $\theta \in (0,1)$, le quantile $q_\theta(Y)$ caractérise la loi de Y.

Le premier quartile, la médiane et le troisième quartile sont respectivement

$$Q_1(Y) = q_{0.25}(Y),\ m(Y) = q_{0.5}(Y) \text{ et } Q_3(Y) = q_{0.75}(Y).$$

Ils représentent les quantiles les plus connus et sont souvent utilisés pour résumer la distribution F de la variable aléatoire univariée Y.
Par exemple, le vecteur $(-0.674, 0, 0.674)^\intercal$ contient les quartiles de la distribution normale standard. Plus généralement, lorsque $Y \sim \mathcal{N}(\mu, \sigma^2)$, nous avons $q_\theta(Y) = \mu + \sigma\Phi^{-1}(\theta)$, où Φ désigne la fonction cumulative de la loi normale centrée réduite.

Il est aussi possible de définir le quantile $q_\theta(Y)$ de niveau θ de la variable Y comme solution du problème de minimisation suivant :

$$\underset{\varsigma \in \mathbb{R}}{\operatorname{argmin}} \mathbb{E}[r_\theta(Y - \varsigma)], \tag{1.1}$$

où $r_\theta(u) = |\theta - \mathbb{1}(u \leq 0)| \cdot |u|$. Il faut noter que le quantile $q_\theta(Y)$ solution de l'équation (1.1) n'est unique que lorsque la fonction de répartition de la variable Y est absolument continue.

Étant donné un échantillon empirique $\{y_1, \cdots, y_n\}$, on peut définir de la même façon le quantile empirique $\widehat{q}_\theta(Y)$ de niveau θ comme solution de l'équation

$$\underset{\varsigma \in \mathbb{R}}{\operatorname{argmin}} \int r_\theta(y - \varsigma) dF_n(y) = \underset{\varsigma \in \mathbb{R}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} r_\theta(y_i - \varsigma) \right\},$$

où $F_n(.)$ est la fonction de répartition empirique de Y.

1.1.2 Expectiles

L'expectile de niveau $\tau \in [0, 1]$ de la variable aléatoire Y est quant à lui défini par :

$$\mu_\tau(Y) = \underset{\varsigma \in \mathbb{R}}{\operatorname{argmin}} \mathbb{E}[\rho_\tau(Y - \varsigma)], \tag{1.2}$$

où $\rho_\tau(u) = |\tau - \mathbb{1}(u \leq 0)| \cdot u^2$. Ainsi, l'expectile d'une variable aléatoire est défini selon la fonction de perte l_2, tandis que le quantile s'obtient avec la fonction de perte l_1.

Newey et Powell (1987) montrent que la solution du problème (1.2) est le paramètre μ_τ solution de l'équation :

$$\mu_\tau(Y) = \mu - \frac{1-2\tau}{1-\tau}\,\mathbb{E}\left[\{Y - \mu_\tau(Y)\}\mathbb{1}\{Y > \mu_\tau(Y)\}\right], \tag{1.3}$$

où $\mu = \mu_{1/2} = \mathbb{E}(Y)$. Cette définition montre que l'expectile est une fonction de l'espérance de la queue de la distribution de Y. Par exemple, si $Y \sim \mathcal{N}(\mu, \sigma^2)$ alors $\mu_\tau(Y) = \mu_\tau$ est solution de l'équation :

$$(2\tau - 1)\mu_\tau\Phi(\xi) + \mu = \tau\mu_\tau + (1-\tau)\lambda(-\xi) - \tau\lambda(\xi),$$

où $\xi = (\mu_\tau - \mu)/\sigma$, $\lambda(x) = \phi(x)/(1 - \Phi(x))$ est la fonction de risque (hazard function), et $\phi(x)$ la densité de la normale standard.

En prenant la condition d'ordre premier de la fonction objectif (1.2), on obtient la relation suivante :

$$\mathbb{E}\left[\psi_\tau(Y - \mu_\tau)(Y - \mu_\tau)\right] = 0, \tag{1.4}$$

où la fonction $\psi_\tau(t) = |\tau - \mathbb{1}(t \leq 0)|$ est la fonction qui assigne des poids τ et $1 - \tau$ respectivement lorsque $Y > \mu_\tau$ et $Y \leq \mu_\tau$. De cette relation (1.4), on déduit une définition plus attrayante de l'expectile :

$$\mu_\tau = \mathbb{E}\left[\frac{\psi_\tau(Y - \mu_\tau)}{\mathbb{E}\left[\psi_\tau(Y - \mu_\tau)\right]}Y\right]. \tag{1.5}$$

Cette dernière définition (1.5), qui est plus pertinente dans le contexte de la régression, révèle que les expectiles, tout comme la moyenne, sont des moyennes pondérées.

La seule particularité est que le poids est une variable aléatoire fonction de la distribution de Y.

Étant donné un échantillon aléatoire, $\{(y_i)\}_{i=1}^n$, l'expectile empirique de niveau τ

$$\widehat{\mu}_\tau = \sum_{i=1}^{n} \frac{\psi_\tau(y_i - \widehat{\mu}_\tau)}{\sum_{i=1}^n \psi_\tau(y_i - \widehat{\mu}_\tau)} y_i$$

est la solution qui minimise la fonction suivante

$$\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_i - \varsigma). \tag{1.6}$$

La plupart des belles propriétés de la fonction expectile sont présentées et démontrées dans l'article original de Newey et Powell (1987). Parmi ces propriétés, deux sont très utiles et regroupées dans le premier théorème de l'article (Newey et Powell, 1987). La première est le fait que la fonction expectile caractérise la fonction de répartition de Y, de la même manière que la fonction quantile. En particulier, à chaque valeur de la fonction de répartition, on peut associer un et un seul expectile. La seconde est sa propriété d'équivariance. L'expectile est équivariant par changement de localisation et d'échelle (location and scale equivariance), c'est-à-dire, pour tout $t \in \mathbb{R}$, et $s > 0$ nous avons :

$$\mu_\tau(sY + t) = s\mu_\tau(Y) + t.$$

La **Figure** 1.1 présente les fonctions de pertes des quantiles et des expectiles pour trois valeurs de $\tau = \theta = (0.1, 0.5, 0.9)$. On peut voir sur ce graphique la distribution asymmétrique des poids selon le niveau. Par exemple, lorsque $\tau = 0.1$, les obsevations à droite de l'expectile et du quantile correspondant reçoivent un faible poids (0.1) et les observations à gauche reçoivent un poids élevé, 0.9.

Relation entre quantile et expectile

Un fait intéressant est que plusieurs caractéristiques et expressions existantes pour les quantiles ont leur analogue avec les expectiles. Par exemple, les quantiles et les expectiles sont des statistiques équivalentes pour des distributions absolument continues. C'est-à-dire qu'à chaque quantile q_θ, de niveau θ, on peut trouver un expectile μ_{τ_θ}, de niveau τ_θ, tel que : $q_\theta = \mu_{\tau_\theta}$. Pour le montrer, considérons, d'un côté, le niveau θ du quantile q_θ défini par :

$$\theta = F(q_\theta) = \mathbb{E}[\mathbb{1}(Y < q_\theta)]. \tag{1.7}$$

D'un autre côté, la condition d'ordre première de la fonction objectif (1.2) sous sa forme intégrale :

$$\int_{\mathbb{R}} \psi_\tau(y - \mu_\tau)(y - \mu_\tau) dF_Y = \tau \int_{\mathbb{R}} |y - \mu_\tau| dF_Y - \int_{-\infty}^{\mu_\tau} (y - \mu_\tau) dF_Y = 0,$$

nous permet de définir le niveau

$$\tau = \frac{\mathbb{E}\left[|Y - \mu_\tau| \mathbb{1}(Y < \mu_\tau)\right]}{\mathbb{E}[|Y - \mu_\tau|]} \tag{1.8}$$

de l'expectile μ_τ.

Maintenant, soit le quantile q_θ de niveau θ. Pour chaque θ fixé, définissons le niveau τ_θ tel que

$$q_\theta = \mu_{\tau_\theta}. \tag{1.9}$$

À partir des expressions (1.7-1.9), nous avons :

$$\tau_\theta = \frac{\theta q_\theta - \mathbb{E}[Y \mathbb{1}(Y < q_\theta)]}{\mathbb{E}[|Y - q_\theta|]} \tag{1.10}$$

et

$$\begin{aligned}\mathbb{E}[|Y-q_\theta|] &= -\int_{-\infty}^{q_\theta}(y-q_\theta)dF_Y + \int_{q_\theta}^{+\infty}(y-q_\theta)dF_Y \\ &= \mu - 2\,\mathbb{E}[Y\mathbb{1}(Y<q_\theta)] - q_\theta(1-2\theta).\end{aligned} \tag{1.11}$$

Ainsi, pour chaque θ fixé, il existe un τ_θ tel que $\mu_{\tau_\theta} = q_\theta$ avec :

$$\tau_\theta = \frac{\theta q_\theta - \mathbb{E}[Y\mathbb{1}(Y<q_\theta)]}{\mu - 2\,\mathbb{E}[Y\mathbb{1}(Y<q_\theta)] - q_\theta(1-2\theta)}.$$

Ce résultat est également vrai pour la classe des modèles conditionnels (ex : régression), comme l'ont montré Yao et Tong (1996). Ainsi, par cette correspondance bijective les expectiles peuvent s'interpréter comme des quantiles et vice versa. Ce résultat est très utile dans des situations où l'estimation des quantiles est très longue et que son unicité n'est pas garantie. D'ailleurs, Yao et Tong (1996) utilisent cette correspondance pour estimer, dans un cadre non-paramétrique, les quantiles conditionnels à partir des expectiles conditionnels. Il existe un développement similaire, avec une notation différente, dans l'article (Waltrup *et al.*, 2015b).

En plus de montrer que l'expectile d'une distribution F est un quantile d'une distribution G, Abdous et Remillard (1995) ont étudié d'autres conditions où les expectiles et les quantiles coïncident. L'expectile peut également s'interpréter comme l'espérance de la queue de la distribution «tail expectation»,

Pour terminer, Koenker (1993) a montré que lorsque la fonction de répartition de Y est définie par

$$F(y) = 1/2\{1 + \text{sign}(y)\sqrt{1+4/(4+y^2)}\}, \quad y \in \mathbb{R}$$

alors

$$q_\theta = \mu_\theta.$$

1.2 Régression Quantile (RQ) et Régression Expectile (RE)

Auparavant, la modélisation de la complexité de la relation des régresseurs et de la variable réponse se traduisait par l'estimation de l'effet des régresseurs sur la médiane ou la moyenne de la variable réponse. Aujourd'hui, avec l'introduction de la régression quantile (RQ) et de la régression expectile (RE), la modélisation permet de capturer l'hétérogénéité de cette relation complexe. Grâce à ces deux outils, la modélisation offre une image globale et détaillée de l'influence des régresseurs, non seulement sur les statistiques centrales, mais sur toute la distribution de la variable réponse. Dans cette section, nous introduisons ces deux nouveaux outils de modélisation dans le cadre d'un modèle linéaire classique.

Considérons l'échantillon $\{(y_i, \boldsymbol{x}_i), i = 1, \ldots, n\}$ de taille n généré par le modèle linéaire classique suivant :

$$y_i = \boldsymbol{x}_i{}^{\mathsf{T}}\boldsymbol{\beta} + \varepsilon_i, \tag{1.12}$$

où $\boldsymbol{x}_i = (x_i^1, \ldots, x_i^p)^{\mathsf{T}} \in \mathbb{R}^p$ est le vecteur des p régresseurs qui influencent la variation de la variable réponse y_i. Le vecteur $\boldsymbol{\beta} \in \mathbb{R}^p$ est le paramètre inconnu qui capture l'effet et la magnitude de cette influence et qu'on doit estimer. Le scalaire ε_i est le terme d'erreur.

1.2.1 Régression Quantile

Le quantile conditionnel $q_\theta(y_i|\boldsymbol{x}_i)$ de niveau θ est défini par :

$$q_\theta(y_i|\boldsymbol{x}_i) = F^{-1}_{y_i|\boldsymbol{x}_i}(\theta) = \inf\{y; F_{y_i|\boldsymbol{x}_i}(y) \geq \theta\},$$

où $F_{y_i|\boldsymbol{x}_i}$ est la fonction de répartition conditionnelle de $y_i|\boldsymbol{x}_i$. Dans le cadre de la régression linéaire classique (1.12), le quantile conditionnel est supposé être une fonction linéaire des régresseurs, c'est-à-dire

$$q_\theta(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i{}^\top \boldsymbol{\beta}_\theta. \tag{1.13}$$

La définition classique du quantile conditionnel, à l'aide de la fonction de répartition de la variable réponse, ne permet pas l'inférence des régresseurs sur les quantiles de la variable réponse. Il a fallu attendre l'introduction, par Koenker et Bassett (1978), du quantile comme solution d'un problème d'optimisation pour pouvoir faire de l'inférence sur les quantiles conditionnels de la distribution de la variable réponse en fonction des régresseurs.

Sous cette caractérisation, Koenker et Bassett (1978) ont défini l'estimateur du paramètre $\boldsymbol{\beta}_\theta \in \mathbb{R}^p$ du modèle de RQ (1.13) comme solution du problème suivant :

$$\widehat{\boldsymbol{\beta}}_\theta \in \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} r_\theta(y_i - \boldsymbol{x}_i{}^\top \boldsymbol{\beta}_\theta) \right\}. \tag{1.14}$$

La solution du problème d'optimisation (1.14) n'est pas analytique car la fonction objectif associée à la RQ n'est pas différentiable à l'origine. Les solutions proposées sont numériques et sont générées par différents algorithmes comme l'algorithme du simplexe, l'algorithme du point intérieur (Koenker, 2005), l'algorithme Majorize-Minimize (MM) (Hunter et Lange, 2000), ou l'approximation linéaire par morceaux (Li *et al.*, 2011). Avec l'essor des données de hautes dimensions, les algorithmes susmentionnés deviennent vite coûteux en terme de calcul. Récemment, des alternatives visant à approximer la fonction de perte par des fonctions lisses ont vu le jour. L'article de Mkhadri *et al.* (2017) donne plus de détails à ce sujet.

En plus d'introduire l'estimateur de la RQ, Koenker et Bassett (1978) ont dérivé ses propriétés asymptotiques. Sous certaines conditions de régularité, l'estimateur de la RQ, $\widehat{\boldsymbol{\beta}}_\theta$, est convergent et asymptotiquement gaussien. Nous avons :

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_\theta - \boldsymbol{\beta}_\theta) \sim \mathcal{N}\Big(\mathbf{0}, \theta(1-\theta)[\boldsymbol{D}_1^Q]^{-1}(\theta)\boldsymbol{D}_0^Q(\theta)[\boldsymbol{D}_1^Q]^{-1}(\theta)\Big),$$

où $\boldsymbol{D}_0^Q(\theta) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\mathsf{T}$ et $\boldsymbol{D}_1^Q(\theta) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n f(q_\theta(y_i|\boldsymbol{x}_i))\boldsymbol{x}_i\boldsymbol{x}_i^\mathsf{T}$, (Koenker, 2005). La fonction f est la densité de y_i et $q_\theta(y_i|\boldsymbol{x}_i)$ est le quantile conditionnel de niveau θ.

La variance de l'estimateur de la RQ dépend de l'estimation de la densité au point $q_\theta(y_i|\boldsymbol{x}_i)$. L'estimation de cette dernière est difficile et coûteuse en temps et pose des problèmes computationnels, (Yin et Cai, 2005). Face à ce problème, la précision des estimateurs de la RQ est souvent obtenue par d'autres approches comme la méthode du Bootstrap. Ce problème a largement été documenté dans la littérature. Pour plus de détails, consulter (Koenker, 2005; Kocherginsky *et al.*, 2005).

Les estimateurs de la RQ ont plusieurs avantages. En plus d'être des estimateurs robustes, ils bénéficient de belles propriétés d'équivariance qui sont très attrayantes pour réduire le temps de calcul des algorithmes. Par exemple, pour toute fonction croissante h, nous avons $q_\theta(h(y)) = h(q_\theta(y))$. En particulier, $q_\theta(sY+t) = sq_\theta(Y)+t$, où $(s,t) \in \mathbb{R}^+ \times \mathbb{R}$. Nous avons les propriétés suivantes (Koenker et Bassett, 1978) :

- $\widehat{\boldsymbol{\beta}}_\theta(\lambda\boldsymbol{y}, \boldsymbol{X}) = \lambda\widehat{\boldsymbol{\beta}}_\theta(\boldsymbol{y}, \boldsymbol{X}), \quad \lambda \in [0, \infty)$;
- $\widehat{\boldsymbol{\beta}}_{1-\theta}(\lambda\boldsymbol{y}, \boldsymbol{X}) = \lambda\widehat{\boldsymbol{\beta}}_\theta(\boldsymbol{y}, \boldsymbol{X}), \quad \lambda \in (-\infty, 0]$;
- $\widehat{\boldsymbol{\beta}}_\theta(\boldsymbol{y} + \boldsymbol{X}\gamma, \boldsymbol{X}) = \widehat{\boldsymbol{\beta}}_\theta(\boldsymbol{y}, \boldsymbol{X}) + \gamma, \quad \gamma \in \mathbb{R}^p$;
- $\widehat{\boldsymbol{\beta}}_\theta(\boldsymbol{y}, \boldsymbol{X}\boldsymbol{A}) = \boldsymbol{A}^{-1}\widehat{\boldsymbol{\beta}}_\theta(\boldsymbol{y}, \boldsymbol{X})$,

où $\boldsymbol{X} = (\boldsymbol{x}_1^\mathsf{T}, \ldots, \boldsymbol{x}_p^\mathsf{T})^\mathsf{T}$ et $\boldsymbol{A}$ est une matrice non-singulière de taille $p \times p$.

Pour terminer, mentionnons que le problème de l'équation (1.14) peut se résoudre comme un estimateur de maximum de vraisemblance du paramètre $\boldsymbol{\beta}_\theta$ lorsque le terme d'erreur, $\epsilon_{i\theta} = y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_\theta$, suit une distribution asymétrique de Laplace (Geraci et Bottai, 2007, 2014).

1.2.2 Régression Expectile

Le modèle de la RE introduit selon Newey et Powell (1987) pour le modèle linéaire classique (1.12) est défini par :

$$\mu_\tau(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau, \quad \text{avec} \quad \mu_\tau(\varepsilon_{i\tau}) = 0. \tag{1.15}$$

Le paramètre $\boldsymbol{\beta}_\tau$ capture l'effet des régresseurs et permet d'étudier leur influence sur la localisation, l'échelle et la forme de la distribution conditionnelle de la variable réponse. L'hypothèse $\mu_\tau(\varepsilon_{i\tau}) = 0$ garantit que l'erreur du modèle est centrée au niveau de son expectile de niveau τ.

Sous l'hypothèse d'homoscédasticité avec un seul régresseur, c'est-à-dire sous le modèle $y = \beta_0 + x\beta_1 + \varepsilon$, l'expectile de la variable réponse est : $\mu_\tau(y|x) = \beta_{0\tau} + x\beta_1$, où $\beta_{0\tau} = \beta_0 + \mu_\tau(\varepsilon_\tau)$. De sorte que, seul l'ordonnée à l'origine est fonction de τ et les fonctions expectiles sont des lignes parallèles. Tandis que, sous l'hypothèse d'hétéroscédasticité, par exemple sous le modèle $y = \beta_0 + x\beta_1 + (\gamma_0 + \gamma_1 x)\varepsilon$, l'expectile de y est donné par : $\mu_\tau(y|x) = \beta_{0\tau} + x\beta_{1\tau}$. Cette fois-ci, l'ordonnée à l'origine, $\beta_{0\tau} = \beta_0 + \gamma_0\mu_\tau(\varepsilon_\tau)$, et la pente, $\beta_{1\tau} = \beta_1 + \gamma_1 x\mu_\tau(\varepsilon_\tau)$, sont fonction du paramètre τ. Ces remarques s'appliquent également au modèle de la RQ.

L'estimateur du paramètre du modèle de la RE (1.15) est défini comme l'unique

solution qui minimise la fonction objectif suivante :

$$\frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i - \boldsymbol{x}_i{}^\mathsf{T}\boldsymbol{\beta}_\tau), \tag{1.16}$$

pour tous les $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$. Comme la fonction de perte associée à la fonction expectile est continûment différentiable, l'application des conditions du premier ordre à l'équation (1.16) génère l'estimateur du modèle de la RE

$$\widehat{\boldsymbol{\beta}}_\tau = \Big(\sum_{i=1}^{n}\boldsymbol{x}_i{}^\mathsf{T}\psi_\tau(\widehat{\varepsilon}_{i\tau})\boldsymbol{x}_i\Big)^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\psi_\tau(\widehat{\varepsilon}_{i\tau})y_i, \tag{1.17}$$

où $\widehat{\varepsilon}_{i\tau} = y_i - \boldsymbol{x}_i{}^\mathsf{T}\widehat{\boldsymbol{\beta}}_\tau$ et $\psi_\tau(t) = |\tau - \mathbb{1}(t \leq 0)|$. Contrairement à l'estimateur du modèle de la RQ, l'estimateur du modèle de la RE se calcule assez facilement et de manière itérative. L'algorithme des moindres carrés pondérés récursifs le génère en un clin d'oeil comparativement au temps de calcul nécessaire pour obtenir l'estimateur du modèle de la RQ. Newey et Powell (1987) ont dérivé les propriétés asymptotiques de l'estimateur du modèle de la RE pour des erreurs i.i.d. Sous certaines conditions de régularités, ils ont montré que :

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau) \sim \mathcal{N}\Big(\mathbf{0}, \boldsymbol{D}_1^{-1}(\tau)\boldsymbol{D}_0(\tau)\boldsymbol{D}_1^{-1}(\tau)\Big),$$

où $\boldsymbol{D}_0(\tau) = \lim_{n\to\infty}\sum_{i=1}^{n} n^{-1}\operatorname{Var}[\psi_\tau(\varepsilon_{i\tau})\varepsilon_{i\tau}]\boldsymbol{x}_i\boldsymbol{x}_i{}^\mathsf{T}$ et $\boldsymbol{D}_1(\tau) = n^{-1}\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}[\psi_\tau(\varepsilon_{i\tau})]\boldsymbol{x}_i\boldsymbol{x}_i{}^\mathsf{T}$.

Les estimateurs des modèles RQ et RE ont des matrices de variance-covariance asymptotiques similaires qui sont utiles pour l'inférence. Lorsqu'on compare les deux matrices des deux modèles, on observe qu'il est plus facile d'estimer celle du modèle RE. En effet, on peut estimer les deux termes $\operatorname{Var}[\psi_\tau(\varepsilon_{i\tau})\varepsilon_{i\tau}]$ et $\mathbb{E}[\psi_\tau(\varepsilon_{i\tau})]$ par leurs moments empiriques respectifs. À cet effet, Newey et Powell (1987) ont proposé cet estimateur robuste $\widehat{\boldsymbol{D}}_1^{-1}(\tau)\widehat{\boldsymbol{D}}_0(\tau)\widehat{\boldsymbol{D}}_1^{-1}(\tau)$, avec

$$\widehat{\boldsymbol{D}}_0(\tau) = n^{-1}\sum_{i=1}^{n}\psi_\tau^2(\widehat{\varepsilon}_{i\tau})\widehat{\varepsilon}_{i\tau}^2\boldsymbol{x}_i\boldsymbol{x}_i{}^\mathsf{T} \text{ et } \widehat{\boldsymbol{D}}_1(\tau) = n^{-1}\sum_{i=1}^{n}\psi_\tau(\widehat{\varepsilon}_{i\tau})\boldsymbol{x}_i\boldsymbol{x}_i{}^\mathsf{T}.$$

Ils ont montré également que l'estimateur est convergent, c'est-à-dire :

$$\widehat{\boldsymbol{D}}_1^{-1}(\tau)\widehat{\boldsymbol{D}}_0(\tau)\widehat{\boldsymbol{D}}_1^{-1}(\tau) \xrightarrow{p} \boldsymbol{D}_1^{-1}(\tau)\boldsymbol{D}_0(\tau)\boldsymbol{D}_1^{-1}(\tau).$$

Soulignons que l'estimateur du modèle de la RE a été étudié auparavant par Aigner *et al.* (1976) qui l'ont présenté sous la forme paramétrique comme un estimateur de maximum de vraisemblance. Sous le modèle (1.12), supposons une distribution normale asymétrique (DNA) pour la variable d'erreur, $\varepsilon \sim \text{DNA}(\varepsilon; \mu_\tau, \sigma^2, \tau)$, de densité

$$f(\varepsilon; \mu_\tau, \sigma^2, \tau) = \frac{2}{\sqrt{\pi\sigma^2}} \frac{\sqrt{\tau(1-\tau)}}{\sqrt{\tau}+\sqrt{1-\tau}} \exp\left\{ -\rho_\tau\left(\frac{\varepsilon-\mu_\tau}{\sigma}\right)\right\}, \tag{1.18}$$

où $\mu_\tau, \sigma,$ et τ sont respectivement les paramètres de localisation, d'échelle et d'asymétrie. Maintenant substituons $\mu_{i\tau}$ par $\mu_{i\tau} = \boldsymbol{x}_i{}^\mathsf{T}\boldsymbol{\beta}_\tau$ et supposons que les n observations sont indépendantes, alors l'estimateur du modèle de la RE est équivalent au maximum de la fonction de vraisemblance définie comme suit :

$$\mathbb{L}(\boldsymbol{\beta}; \sigma, \tau, \boldsymbol{y}) \propto \sigma^{-2n} \exp\left\{ -\sum_{i=1}^{n} \rho_\tau\left(\frac{y_i - \boldsymbol{x}_i{}^\mathsf{T}\boldsymbol{\beta}_\tau}{\sigma}\right)\right\},$$

où $\boldsymbol{y} = (y_1, \ldots, y_n)^\mathsf{T}$. La distribution asymétrique (DNA) présentée ici (1.18) n'est pas à confondre avec celle présentée par Azzalini (1985). Dans son article, Azzalini introduit une nouvelle classe de densités définie à partir du produit de la densité standard $f_Y(\cdot)$ et de la fonction de répartition $F_Y(\cdot)$ d'une variable aléatoire Y. Cette nouvelle classe est paramétrée par λ et la distribution normale correspond à $\lambda = 0$.

Nous terminons cette section par la présentation des fonctions quantile et expectile de quelques distributions usuelles. Dans la **Figure** 1.2, on peut observer que les deux fonctions, quantile et expectile, sont strictement croissantes et couvrent toutes les valeurs de l'ensemble $I_F = \{y | 0 < F(y) < 1\}$. Les courbes des deux fonctions

s'entrecroisent au point $\theta = \tau = 0.5$ pour les lois symétrique (normale, Student). Enfin, notons la proximité des courbes quantile et expectile de la loi de Student.

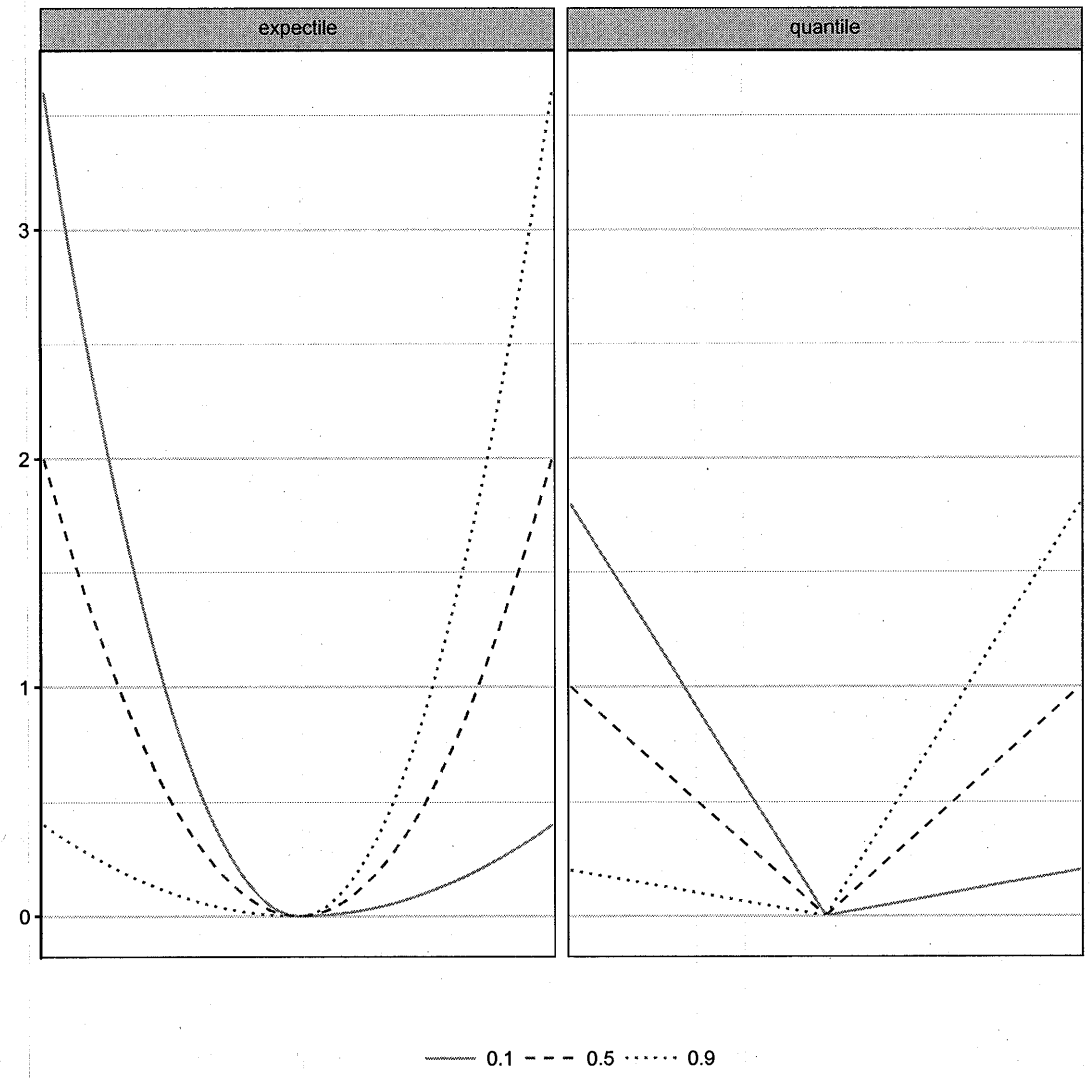

Figure 1.1: Fonctions de perte des expectiles et des quantiles

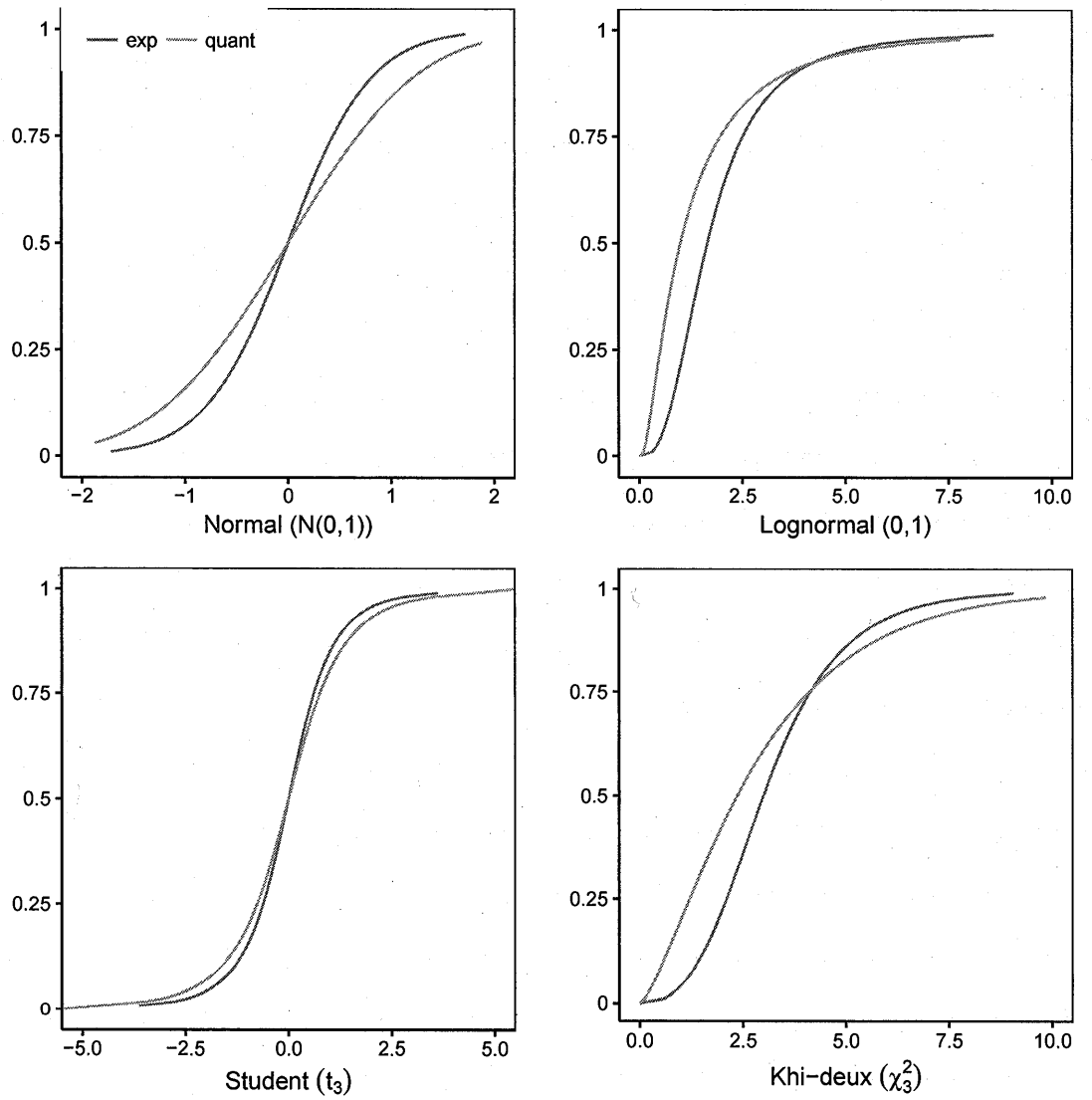

Figure 1.2: Expectiles et quantiles de quelques distributions usuelles

1.3 Analyse des données longitudinales

Dans la présente section nous introduisons les données longitudinales et quelques méthodes d'analyse. Les méthodes d'analyse des données longitudinales se classent en deux grandes familles : les modèles marginaux et les modèles conditionnels. Nous présenterons des méthodes apparentées aux deux.

Les données utilisées en statistique se répartissent en deux classes : les données expérimentales et les données non-expérimentales, observées ou observationnelles.

Les données expérimentales proviennent d'une étude expérimentale, dans laquelle le chercheur possède un certain contrôle sur la construction de l'échantillon. Le chercheur peut assigner selon une méthode aléatoire ou non-aléatoire les sujets dans un groupe traité/exposé ou dans un groupe contrôle/placebo. Les données expérimentales sont considérées comme les données de référence. Les résultats obtenus sont dépourvus d'un certain nombre de biais, entre autres, le biais de sélection. Cela dit, les études expérimentales sont rares en raison de leur coût élevé, et avant tout pour des questions d'éthique. En effet, nous ne pouvons imaginer que pour étudier l'efficacité d'un nouveau médicament un chercheur décide lui-même quels sujets reçoivent le traitement et quels sujets reçoivent le placebo, et ceci, sans leur consentement. Nous ne pouvons pas non plus imaginer que pour déterminer l'effet du tabagisme de la mère sur la mortalité infantile, un chercheur détermine quelles femmes devront fumer ou s'abstenir de fumer pendant la grossesse. Tous ces exemples transgressent l'éthique.

Les autres types de données, non-expérimentales, sont désignées comme des don-

nées observationnelles. Elles constituent la majorité des données empiriques et sont générées par un processus qui échappe au contrôle du chercheur. Parmi ces données observationnelles, les données longitudinales sont de loin les données les plus appréciées.

Les données longitudinales sont recueillies et analysées dans plusieurs champs d'application tels que l'épidémiologie (Smith *et al.*, 2015), la génétique (Furlotte *et al.*, 2012), l'économie (Hsiao, 2007), pour ne citer que ceux-là. Une étude longitudinale consiste à suivre la dynamique d'une cohorte et à recueillir périodiquement les mesures des caractéristiques des sujets de la cohorte sur une période donnée. Une étude longitudinale capture la dynamique, le développement ou le changement de la population et permet de mieux distinguer/expliquer l'hétérogénéité. Les études longitudinales sont appréciées, entre autres, parce qu'elles offrent une meilleure inférence des paramètres du modèle, notamment une plus grande puissance statistique et des prédictions plus précises. Elles sont également très utiles pour séparer l'effet de l'âge de l'effet de la cohorte. Elles sont nécessaires dans l'étude des relations causales avec des données observationnelles. En effet, un des critères de causalité est la règle de temporalité, c'est-à-dire que la cause précède l'effet (Rothman *et al.*, 2008).

Les données longitudinales sont caractérisées par la présence d'une dépendance intra-groupe, c'est-à-dire que les observations provenant d'un même sujet sont inter-dépendantes. Cette structure de dépendance lui confère son principal avantage par rapport aux données transversales. Néanmoins, cette dépendance complique l'analyse statistique et doit être modélisée afin de générer des estimateurs efficaces et non biaisés. Les méthodes standards de modélisation, comme la régression linéaire simple, produisent des estimateurs biaisés de la matrice de covariance lorsqu'elles sont ajustées aux don-

nées longitudinales. Pour tenir compte de cette structure de dépendance, plusieurs méthodes ont été élaborées. Elles se classent en deux grandes familles : les modèles marginaux et les modèles conditionnels. Ces familles se distinguent principalement par la manière dont elles modélisent la dépendance.

Par la suite, pour simplifier les notations et améliorer la lisibilité, les modèles mathématiques seront présentés avec des données implicitement équilibrées. C'est-à-dire que tous les individus/sujets auront le même nombre d'observations, m. Cela dit, les résultats obtenus se généralisent également aux données déséquilibrées.

1.3.1 Modèles marginaux

Les modèles marginaux modélisent l'effet des régresseurs sur la moyenne de la variable réponse avec une spécification du modèle identique à celle d'un modèle pour l'analyse des données transversales. Les modèles marginaux n'insèrent pas d'effets aléatoires dans la spécification de leur modèle. Dans le cadre des modèles marginaux, la dépendance des données de la variable réponse est caractérisée par la formulation d'une structure de corrélation hypothétique. Pour plus de détails, voir l'excellent livre de Hardin et Hilbe (2012).

Une méthode d'estimation couramment utilisée dans le cadre des modèles marginaux est la méthode d'estimation des équations généralisées, couramment désignée GEE. Le modèle GEE a été proposé par Liang et Zeger (1986) dans le but de prolonger les modèles linéaires généralisés (McCullagh et Nelder, 1989) à l'analyse des données longitudinales. La méthode GEE résume la relation entre les régresseurs et la variable réponse en estimant l'effet sur la moyenne marginale populationnelle. Elle

tient compte de la dépendance intra-groupe des données par l'inclusion formelle d'une matrice de corrélation (des observations de la variable réponse d'un même individu) dite *working correlation* dans l'équation d'estimation des paramètres du modèle.

Pour mieux formaliser le problème, considérons les données $\{y_{ij}, \boldsymbol{x}_{ij}\}_{1\leq i\leq n, 1\leq j\leq m}$ observées sur n sujets, avec y_{ij} la variable réponse du sujet i mesurée au temps j et $\boldsymbol{x}_{ij} = (x_{ij}^1, \ldots, x_{ij}^p)^\mathsf{T}$ le vecteur des régresseurs de dimension p. Lorsque les données sont regroupées par sujet, nous notons $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im})^\mathsf{T}$ le vecteur réponse de taille m du sujet i et $\boldsymbol{X}_i$ la matrice des régresseurs de dimension $m \times p$.

Le modèle GEE est une méthode de quasi-vraisemblance (Hardin et Hilbe, 2012). Sa spécification se limite à la caractérisation des deux premiers moments, l'espérance marginale et la variance, de la variable réponse et à la supposition d'une structure de corrélation qui capture le mieux la dépendance des observations provenant d'un même sujet/groupe. Dans un premier temps, l'espérance marginale de la variable réponse $\mathbb{E}[y_{ij}|\boldsymbol{x}_{ij}] = \mu_{ij}$ est liée aux régresseurs par une fonction connue $g(\mu_{ij}) = \eta_{ij} = \boldsymbol{x}_{ij}{}^\mathsf{T}\boldsymbol{\beta}$, où $\boldsymbol{\beta} \in \mathbb{R}^p$ représente le vecteur des paramètres. Dans un deuxième temps, il faut exprimer la variance de la variable réponse qui est fonction de l'espérance à travers la fonction de variance $\nu(\cdot)$. Enfin, il faut assumer une structure de corrélation pour modéliser la dépendance des observations de chaque sujet.

L'estimateur $\widehat{\boldsymbol{\beta}}$ du paramètre $\boldsymbol{\beta}$ par la méthode des GEE est obtenu par la résolution du système d'équation $\boldsymbol{S}(\boldsymbol{\beta}) = \mathbf{0}$, où

$$\boldsymbol{S}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{D}_i^{\mathsf{T}} \boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})), \tag{1.19}$$

où $\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ et $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im})^{\mathsf{T}}$. La matrice de variance-covariance de l'estimateur $\widehat{\boldsymbol{\beta}}$ est définie par

$$\boldsymbol{V}_i(\boldsymbol{\alpha}) = \phi \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i(\boldsymbol{\alpha}) \boldsymbol{A}_i^{1/2},$$

où ϕ est le paramètre d'échelle, $\boldsymbol{A}_i = \text{diag}[\nu(\mu_{ij})]_{j=1}^{m}$ et $\boldsymbol{\alpha}$ le vecteur des paramètres de taille q qui caractérise la matrice de corrélation $\boldsymbol{R}_i(\boldsymbol{\alpha})$ associée à la matrice de covariance $\boldsymbol{V}_i(\boldsymbol{\alpha})$.

La matrice de corrélation $\boldsymbol{R}_i(\boldsymbol{\alpha})$ est de dimension $m \times m$. Elle décrit la structure de corrélation des observations provenant d'un même sujet par l'intermédiaire de $\boldsymbol{\alpha}$. Il existe plusieurs types de structures de corrélation. Nous présentons uniquement les plus usuels. Pour une liste exhaustive des autres formes de structure de corrélation, voir le livre de Hardin et Hilbe (2012).

La structure de corrélation indépendante est la forme la plus simple de la matrice de corrélation. La structure de corrélation indépendante suppose une corrélation nulle, $\rho_{ts} = 0, \forall t \neq s$, pour toute paire distincte du même sujet. L'hypothèse n'est pas toujours vérifiée, mais Pan (2001) montre que l'estimateur associé est généralement performant. La structure échangeable est une simple extension de la corrélation indépendante. Elle suppose une corrélation commune, $\rho_{ts} = \alpha, \forall t \neq s$, entre chaque paire de mesures provenant d'un même sujet. Elle est aussi désignée par les termes «corrélation commune» ou «corrélation composée». Le paramètre α est un scalaire

et est le même pour tous les sujets de l'échantillon. La corrélation échangeable est généralement plus appropriée pour des ensembles de données dans lesquels les observations n'ont aucune dépendance au temps. La structure de corrélation AR1 définit la corrélation d'une paire d'observations comme une fonction décroissante de leur distance dans le temps, $\rho_{ts} = \alpha^{|t-s|}$. Cette structure de corrélation attribue la corrélation la plus élevée à la paire d'observations adjacente et la corrélation la plus faible à la paire distante. Ce type de corrélation est indiqué lorsque les mesures répétées ont une dépendance temporelle. Comme pour la corrélation échangeable, le paramètre α est un scalaire. La corrélation non structurée, comme son nom l'indique, n'impose aucune structure à la matrice de corrélation et définit la corrélation de n'importe quelle paire de mesures différemment sans aucun modèle explicite, $\rho_{ts} = \alpha_{ts}, \forall t \neq s$. De ce fait, chaque paire d'observations a une corrélation distincte.

Le système des équations d'estimation généralisées (1.19) dépend du paramètre d'intérêt $\boldsymbol{\beta}$ et du paramètre de corrélation $\boldsymbol{\alpha}$ et il n'a pas de solution explicite. Une procédure itérative d'estimation en deux étapes est nécessaire pour estimer $\boldsymbol{\beta}$ et les paramètres de nuisance $\boldsymbol{\alpha}$ et d'échelle ϕ. Le paramètre ϕ représente le paramètre de sur-dispersion dans les modèles linéaires généralisées (GLM). L'algorithme d'estimation est présenté ci-dessous en supposant une corrélation échangeable.

Algorithme : L'algorithme des équations d'estimation généralisées

---

Étape 1. Initier $\widetilde{\boldsymbol{\beta}}^{(0)} = \widehat{\boldsymbol{\beta}}_I$, l'estimateur des moindres carrés ordinaires.

Étape 2. Pour $r = 1, 2, \ldots$, calculer la mise à jour de l'estimateur à l'étape $r-1$.

$$\widetilde{\phi}^{(r)} \leftarrow \frac{1}{N-p} \sum_{i=1}^{n} \sum_{t=1}^{m} \widetilde{u}_{it}^2,$$

$$\widetilde{\alpha}^{(r)} \leftarrow \frac{1}{(N_1-p)\widetilde{\phi}^{(r)}} \sum_{i=1}^{n} \sum_{t<s}^{m} \widetilde{u}_{it}\widetilde{u}_{is},$$

où $N_1 = \frac{nm(m-1)}{2}$ et $\widetilde{u}_{ij} = y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}}\widetilde{\boldsymbol{\beta}}^{(r-1)}$.

Étape 3. Calculer

$$\widetilde{\boldsymbol{\beta}}^{(r)} \leftarrow \widetilde{\boldsymbol{\beta}}^{(r-1)} + \left[\sum_{i=1}^{n} \boldsymbol{D}_i{}^{\mathsf{T}} \boldsymbol{V}_i^{-1}(\widetilde{\alpha}^{(r-1)}) \boldsymbol{D}_i\right]^{-1} \boldsymbol{S}\left(\widetilde{\boldsymbol{\alpha}}^{(r-1)}, \widetilde{\boldsymbol{\beta}}^{(r-1)}\right).$$

Étape 4. Itérer les **Étapes** 2-3, jusqu'à convergence.

L'algorithme s'applique également aux autres formes de structures de corrélation. Pour l'adapter, il suffit de choisir le paramètre de corrélation approprié $\boldsymbol{\alpha}$, qui est soit un scalaire, soit un vecteur ou une matrice, selon le type de corrélation. Par exemple, pour une structure de corrélation AR1 auto-régressive d'ordre 1, le paramètre scalaire α est mis à jour à l'**Étape** 2 de l'algorithme précédent comme suit :

$$\widetilde{\alpha} \leftarrow \frac{1}{(N_2-p)\widetilde{\phi}} \sum_{i=1}^{n} \sum_{t<m-1} \widetilde{u}_{it}\widetilde{u}_{i,t+1,}, \; N_2 = n(m-1).$$

Pour une structure de corrélation non-structurée, chacun des $m(m+1)/2$ éléments du paramètre $\boldsymbol{\alpha}$ est estimé par

$$\widetilde{\alpha}_{ts} \leftarrow \frac{1}{(N-p)\widetilde{\phi}} \sum_{i=1}^{n} \widetilde{u}_{it}\widetilde{u}_{is}, \ N = nm.$$

Lorsque le modèle est correctement spécifié, Liang et Zeger (1986) ont montré, sous certaines conditions régulières, que l'estimateur du paramètre d'intérêt $\boldsymbol{\beta}$ par la méthode GEE est convergent et asymptotiquement normal. Plus précisément, nous avons

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\Big(\mathbf{0}, \mathrm{Var}(\widehat{\boldsymbol{\beta}})\Big),$$

où

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{B}^{-1}\boldsymbol{M}\boldsymbol{B}^{-1}$$

avec $\boldsymbol{B} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_i^{\mathsf{T}} \boldsymbol{V}_i^{-1} \boldsymbol{D}_i$, $\quad \boldsymbol{M} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_i^{\mathsf{T}} \boldsymbol{V}_i^{-1} \mathrm{Var}(\boldsymbol{y}_i) \boldsymbol{V}_i^{-1} \boldsymbol{D}_i$

et

$$\mathrm{Var}(\boldsymbol{y}_i) = \mathbb{E}[(\boldsymbol{y}_i - \boldsymbol{\mu}_i)(\boldsymbol{y}_i - \boldsymbol{\mu}_i)^{\mathsf{T}}].$$

Voir (Liang et Zeger, 1986) pour plus de détails. Afin d'utiliser ce nouvel estimateur pour faire de l'inférence, comme construire des intervalles de confiance ou réaliser des tests d'hypothèses, il faudra estimer sa matrice de variance-covariance. L'estimateur proposé par Liang et Zeger (1986) est une adaptation de l'estimateur robuste de White (1980), communément appelé l'estimateur «sandwich». Ainsi l'estimateur de la matrice de variance-covariance est :

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{B}}^{-1}\widehat{\boldsymbol{M}}\widehat{\boldsymbol{B}}^{-1},$$

où

$$\widehat{\boldsymbol{B}} = n^{-1} \sum_{i=1}^{n} \widehat{\boldsymbol{D}}_i^{\mathsf{T}} \boldsymbol{V}_i^{-1}(\widehat{\alpha}, \widehat{\phi}) \widehat{\boldsymbol{D}}_i$$

et

$$\widehat{\boldsymbol{M}} = n^{-1} \sum_{i=1}^{n} \widehat{\boldsymbol{D}}_i{}^{\mathsf{T}} \boldsymbol{V}_i^{-1}(\widehat{\alpha}, \widehat{\phi}) \Big[\boldsymbol{y}_i - \boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}) \Big] \Big[\boldsymbol{y}_i - \boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}) \Big]^{\mathsf{T}} \boldsymbol{V}_i^{-1}(\widehat{\alpha}, \widehat{\phi}) \widehat{\boldsymbol{D}}_i.$$

Notez que lorsque la matrice de corrélation $\boldsymbol{R}_i$ correspond à la vraie structure de corrélation, alors l'estimateur de la variance se réduit à

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}) = n^{-1} \sum_{i=1}^{n} \widehat{\boldsymbol{D}}_i{}^{\mathsf{T}} \boldsymbol{V}_i^{-1}(\widehat{\alpha}, \widehat{\phi}) \widehat{\boldsymbol{D}}_i.$$

L'approche GEE propose une multitude de structures de corrélation pour modéliser la dépendance intra-groupe. Face à toutes ces possibilités, le processus de sélection de la structure de corrélation peut s'avérer complexe et difficile. Heureusement, il y a des critères qui ont été développés pour aider dans cet exercice, (Jang, 2011). Dans la plupart des logiciels statistiques le critère de quasi-vraisemblance «quasi-likelihood information criterion (QIC)» est implémenté pour choisir parmi les différentes structures de corrélation. Le QIC, développé par Pan (2001), permet simultanément la sélection du modèle et la sélection de la structure de corrélation. Le QIC, une adaptation du critère (AIC) pour le modèle GEE, est défini par :

$$QIC(\boldsymbol{R}) = -2Q_u(\widehat{\boldsymbol{\beta}}(\boldsymbol{R}), \widehat{\phi}) + 2\,\mathrm{Trace}\left(\widehat{\boldsymbol{\Omega}}_{\mathbf{I}} \widehat{\boldsymbol{V}}_{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}})\right)$$

où $\widehat{\boldsymbol{\beta}}(\boldsymbol{R})$ est l'estimateur GEE avec la corrélation d'intérêt $\boldsymbol{R}$. La matrice $\widehat{\boldsymbol{V}}_{\boldsymbol{R}}$ est l'estimateur robuste de la matrice de covariance et $\widehat{\boldsymbol{\Omega}}_{\mathbf{I}}$ est l'inverse de l'estimation de la covariance du modèle GEE sous l'hypothèse de la structure d'indépendance évaluée à $\widehat{\boldsymbol{\beta}}(\boldsymbol{R})$. La fonction Q_u est la fonction quasi-vraisemblance sous l'hypothèse de la structure d'indépendance évaluée à $\widehat{\boldsymbol{\beta}}(\boldsymbol{R})$. Elle a pour expression :

$$Q_u(\widehat{\boldsymbol{\beta}}(\boldsymbol{R}), \phi) = \sum_{i=1}^{n} \sum_{j=1}^{m} Q(\widehat{\boldsymbol{\beta}}(\boldsymbol{R}), \widehat{\phi}; (y_{ij}, \boldsymbol{x}_{ij})),$$

où $Q(\mu, \widehat{\phi}; y) = \int_y^u [(y-t)/\widehat{\phi} V(t)] dt$ est défini pour différents modèles par (McCullagh et Nelder, 1989; Hardin et Hilbe, 2012).

Même si le critère QIC de Pan est utile dans le processus de sélection de la structure de corrélation qui modélise le mieux la dépendance intra-groupe, il faut se fier, lorsque c'est disponible, aux connaissances de la structure de corrélation provenant d'autres sources.

Dans la section suivante nous présentons la deuxième famille des modèles conditionnels.

1.3.2 Modèles conditionnels

Les modèles conditionnels se distinguent des modèles marginaux par l'introduction d'un paramètre individuel dans l'équation du modèle. Ce paramètre individuel qu'on désigne généralement par «effets aléatoires» permet de tenir compte de la dépendance entre les observations provenant d'un même sujet. Plusieurs théories relative à cette famille ont été développées séparément dans divers domaines d'application. De ce fait, les méthodes élaborées pour étudier les modèles de cette famille se nomment différemment selon le champ d'application. En sciences de la santé, comme en biostatistique ou en épidémiologie, les termes modèles mixtes linéaires généralisés (GLMM) ou modèles multiniveaux ou hiérarchiques sont fréquents. Tandis qu'en économétrie, les méthodes pour ajuster ces modèles avec effets aléatoires sont réparties en deux sous-groupes : les modèles avec effets fixes et les modèles avec effets aléatoires.

D'abord, dressons le portrait. Considérons le modèle linéaire général employé pour l'analyse des données longitudinales avec un modèle conditionnel :

$$y_{ij} = \boldsymbol{x}_{ij}{}^{\top}\boldsymbol{\beta} + \boldsymbol{g}_i{}^{\top}\boldsymbol{\gamma} + \zeta_i + u_{ij},\ i = 1, \ldots, n;\ j = 1, \ldots, m, \tag{1.20}$$

où y_{ij} est la variable réponse, $\boldsymbol{x}_{ij} = (x_{ij}^1, x_{ij}^2, \ldots, x_{ij}^p)^\mathsf{T} \in \mathbb{R}^p$ est le vecteur des régresseurs mesurés sur l'individu i au temps j. Le vecteur $\boldsymbol{g}_i \in \mathbb{R}^q$ est le vecteur des caractéristiques individuelles observées telles que : le sexe et le statut marital et $\boldsymbol{\gamma} \in \mathbb{R}^q$ le vecteur associé. Le scalaire ζ_i représente les caractéristiques individuelles non-observées comme la dextérité, les facteurs génétiques et environnementaux et le scalaire u_{ij} est le terme d'erreur. Le vecteur $\boldsymbol{\beta} \in \mathbb{R}^p$ est le paramètre d'intérêt que nous voulons estimer.

À cette étape, nous présenterons les deux modèles : modèle avec effets fixes et modèle avec effets aléatoires. Malgré leur popularité en économétrie, les termes «effets fixes» et «effets aléatoires» portent souvent confusion. À première vue, le mot«effets fixes» suggère que le paramètre individuel dans le modèle avec«effets fixes» est fixe et que le paramètre individuel dans le modèle avec «effets aléatoires» est aléatoire. Ce n'est pas le cas. Dans les deux modèles, le paramètre individuel est un paramètre aléatoire. Les termes utilisés pour désigner les deux modèles font référence à la relation ou à la nature de l'association qui existe entre les régresseurs $\boldsymbol{x}_{ij}$ et les effets individuels non-observés ζ_i, (Greene, 2011).

Modèle avec effets fixes

Le modèle avec effets fixes (EF) est un modèle communément utilisé pour l'analyse des données longitudinales. Dans sa spécification, le modèle EF prend en compte les régresseurs non mesurés ou non mesurables et inclus dans l'équation du modèle. De plus, sa spécification sous-tend la possibilité d'une corrélation entre les régresseurs omis et ceux présents dans le modèle. Cette dernière caractéristique fait du modèle EF un outil attrayant pour l'analyse de données longitudinales. Par exemple, dans

l'estimation de l'effet du rendement scolaire, le niveau d'éducation est corrélé à l'aptitude individuelle qui est une caractéristique individuelle souvent omise du modèle. Dans les études épidémiologiques et biologiques, le facteur génétique ou environnemental des sujets est indisponible alors qu'il est une caractéristique individuelle corrélée à la variable d'intérêt du modèle. Dans la suite, nous présentons formellement le modèle EF.

Considérons le cadre général représenté par l'équation (1.20), le modèle EF suppose l'existence d'une corrélation entre les effets individuels ζ_i, et les régresseurs $\boldsymbol{x}_{ij}$. Cette relation est formalisée par l'équation suivante :

$$\mathbb{E}[\zeta_i|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im}] = h(\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im}) = \eta_i, \tag{1.21}$$

où h est une fonction non-définie. Contrairement à l'intuition, le terme«effets fixes» du modèle EF désigne le fait que la dépendance entre les régresseurs et les effets individuels est fixée à η_i (1.21). L'inclusion de cette hypothèse dans l'équation du modèle général (1.20) permet de définir l'équation du modèle EF sous la forme suivante :

$$\begin{aligned} y_{ij} &= \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{g}_i{}^{\mathsf{T}}\boldsymbol{\gamma} + \eta_i + u_{ij} + [\zeta_i - \eta_i] \\ &= \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{g}_i{}^{\mathsf{T}}\boldsymbol{\gamma} + \eta_i + \varepsilon_{ij} \\ &= \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta} + \eta_i + \varepsilon_{ij}, \qquad (\boldsymbol{g}_i{}^{\mathsf{T}}\boldsymbol{\gamma} \text{ est inclus dans } \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta}), \end{aligned} \tag{1.22}$$

où $\varepsilon_{ij} = u_{ij} + [\zeta_i - \eta_i]$. Par construction le terme $\zeta_i - \eta_i$ est non-corrélé aux régresseurs et $\mathbb{E}[\zeta_i - \eta_i] = 0$. Avec cette transformation, la moyenne conditionnelle de y_{ij} est modélisée par :

$$\mathbb{E}[y_{ij}|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im}] = \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta} + \eta_i.$$

Pour des raisons pratiques, on représente souvent l'équation du modèle EF (1.22) sous la forme matricielle suivante :

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{1.23}$$

où $\boldsymbol{y}$ est le $N \times 1$ vecteur, $\boldsymbol{X}$ est la matrice $N \times p$, $\boldsymbol{\eta}$ et $\boldsymbol{\varepsilon}$ sont respectivement les vecteurs de taille $n \times 1$ et $N \times 1$, avec $N = nm$. La matrice $\boldsymbol{Z}$ de taille $N \times n$ est connue sous le nom de matrice d'incidence et permet de tenir compte de la dépendance des données.

Les estimateurs des paramètres $\boldsymbol{\beta}$ et $\boldsymbol{\eta}$ du modèle EF sont les vecteurs qui minimisent la fonction de risque suivante :

$$R_{EF}(\boldsymbol{\beta}, \boldsymbol{\eta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\eta}\|^2, \tag{1.24}$$

où la norme est définie par $\|\boldsymbol{u}\|^2 = \boldsymbol{u}^\top \boldsymbol{u}$. Dans le cadre du modèle EF, en plus du paramètre d'intérêt $\boldsymbol{\beta}$, il faut estimer le paramètre individuel $\boldsymbol{\eta}$ non-observé dont la dimension augmente en fonction de la taille de l'échantillon n. La nature infinie de la dimension de ce paramètre individuel introduit des complications, communément désignées par : *problème du paramètre d'incidence.* Il existe plusieurs méthodes d'estimation pour résoudre ce problème d'incidence, entre autres, le «first difference», le «least square dummy variable» ou la méthode de la transformation intra-groupe («within-transformation»). Toutes ces méthodes produisent le même estimateur. Cependant, la méthode de la transformation intra-groupe est la plus pratique computationnellement. Elle permet une estimation directe du paramètre d'intérêt $\boldsymbol{\beta}$ indépendamment du paramètre individuel $\boldsymbol{\eta}$, c'est-à-dire sans pour autant devoir estimer cette dernière.

La méthode intra-groupe ou «within-transformation» s'exécute en deux étapes. Dans un premier temps, on applique une transformation linéaire (matrice de projection) à l'équation initiale du modèle EF (1.22) pour éliminer le paramètre individuel. Dans un second temps, on ajuste la méthode des moindres carrés ordinaires (MCO) sur le

nouveau modèle. L'estimateur de la méthode intra-groupe ou «within-transformation» du modèle EF $\widehat{\boldsymbol{\beta}}_W$ qui en découle est défini par :

$$\widehat{\boldsymbol{\beta}}_W = (\boldsymbol{X}^\top \boldsymbol{M}_{\boldsymbol{Z}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{M}_{\boldsymbol{Z}} \boldsymbol{y}, \tag{1.25}$$

où les matrices de projections sont définies par :

$$\boldsymbol{M}_{\boldsymbol{Z}} = \mathbb{I}_N - \boldsymbol{P}_{\boldsymbol{Z}} \text{ et } \boldsymbol{P}_{\boldsymbol{Z}} = \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top.$$

Notons que, par construction, la matrice $\boldsymbol{M}_{\boldsymbol{Z}}$ est orthogonale aux régresseurs invariants dans le temps. Ainsi, le modèle EF ne permet pas l'inférence des régresseurs invariants dans le temps. Cela constitue la principale limite du modèle EF. Cela dit, il y a différentes astuces pour éluder ce problème, comme la multiplication par la variable temps des régresseurs invariants.

Il a été démontré, sous certaines conditions de régularité, que l'estimateur intra-groupe est sans biais, convergent et normalement distribué asymptotiquement (Hsiao, 2015; Arellano, 2003). En particulier, ces propriétés asymptotiques sont prouvées lorsque la taille de l'échantillon n, ou le nombre de période m, ou les deux tendent vers l'infini. Si l'intérêt porte sur le paramètre d'incidence, on peut l'estimer par :

$$\widehat{\eta}_i = \overline{y}_i - \overline{\boldsymbol{x}}_i^\top \widehat{\boldsymbol{\beta}}_W, \; i = 1, \ldots, n,$$

où $\overline{y}_i = m^{-1} \sum_{k=1}^m y_{ik}$ et $\overline{\boldsymbol{x}}_i = (\overline{x}_i^1, \ldots, \overline{x}_i^p)^\top$ avec $\overline{x}_i^j = m^{-1} \sum_{k=1}^m x_{ik}^j$. Il faut noter que cet estimateur est non-convergent lorsque le nombre de périodes est fixé.

La matrice de variance-covariance asymptotique de l'estimateur intra-groupe $\widehat{\boldsymbol{\beta}}_W$

est donnée par

$$\begin{aligned}\mathrm{Var}(\widehat{\boldsymbol{\beta}}_W) &= \lim_{n\to\infty} n\Big[\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{X}^*\Big]^{-1}\boldsymbol{X}^{*\mathsf{T}}\,\mathrm{Var}(\boldsymbol{\varepsilon}^*)\boldsymbol{X}^*\Big[\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{X}^*\Big]^{-1} \\ &= \lim_{n\to\infty} n\Big[\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{X}_i^*\Big]^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\,\mathbb{E}(\boldsymbol{\varepsilon}_i^*\boldsymbol{\varepsilon}_i^{*\mathsf{T}})\boldsymbol{X}_i^*\Big[\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{X}_i^*\Big]^{-1},\end{aligned}$$

où $\boldsymbol{X}^* = \boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{X}$ et $\boldsymbol{\varepsilon}^* = \boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{\varepsilon}$.

Afin de produire une inférence statistique valide, Arellano (1987) a proposé un estimateur robuste qui tient compte de l'hétérogénéité des données et de la corrélation des observations provenant d'un même individu/sujet. Cet estimateur est donné par :

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}_W) = n\Big[\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{X}_i^*\Big]^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\widehat{\boldsymbol{\varepsilon}}^*{}_i\widehat{\boldsymbol{\varepsilon}}^*{}_i{}^{\mathsf{T}}\boldsymbol{X}_i^*\Big[\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{X}_i^*\Big]^{-1},$$

où $\widehat{\boldsymbol{\varepsilon}}_i^* = \boldsymbol{y}_i^* - \boldsymbol{X}_i^*\widehat{\boldsymbol{\beta}}_W$.

Il est également possible d'estimer le modèle EF par l'approche de la vraisemblance, voir par exemple (Hsiao, 2015; Arellano, 2003).

Modèle avec effets aléatoires

L'autre méthode communément utilisée pour modéliser les données longitudinales est le modèle avec effets aléatoires (EA). Le modèle EA se distingue du modèle EF par son hypothèse d'indépendance sur la relation entre les effets individuels non-observables et les régresseurs du modèle (Cameron et Trivedi, 2005). Cela dit, cette hypothèse n'est pas toujours vérifiée dans les études empiriques. Dans la réalité, il y a toujours des caractéristiques individuelles qui ne sont pas mesurées et qui se retrouvent dans le terme d'erreur du modèle.

L'hypothèse d'indépendance entre les effets individuels et les régresseurs est représentée par la formule suivante :

$$\mathbb{E}[\zeta_i|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im}] = \eta. \tag{1.26}$$

L'intégration de cette hypothèse d'indépendance permet de réécrire l'équation générale du modèle (1.20) et de définir le modèle EA comme suit :

$$\begin{aligned} y_{ij} &= \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{g}_i{}^{\mathsf{T}}\boldsymbol{\gamma} + \eta + \{\zeta_i - \eta\} + u_{ij} \\ &= \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{g}_i{}^{\mathsf{T}}\boldsymbol{\gamma} + \eta + v_i + u_{ij} \\ &= \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta} + v_i + u_{ij}, \qquad (\boldsymbol{g}_i{}^{\mathsf{T}}\boldsymbol{\gamma} \text{ et } \eta \text{ sont inclus dans } \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta}) \end{aligned} \tag{1.27}$$

où η est une constante et v_i est l'effet aléatoire du groupe i. Le vecteur $\boldsymbol{x}_{ij}$ inclut la constante et toute autre variable invariante dans le temps. Avec cette transformation, nous avons $\mathbb{E}[v_i|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im}] = \mathbb{E}[\zeta_i - \eta|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im}] = 0$ et la moyenne conditionnelle de y_{ij} est modélisée par :

$$\mathbb{E}[y_{ij}|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im}] = \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta}.$$

Ainsi, la moyenne conditionnelle du modèle EA est similaire à celle d'un modèle de la famille marginale, bien que nous ayons utilisé un modèle de la famille conditionnelle avec la présence d'un effet aléatoire ζ_i dans l'équation initiale du modèle (1.20). Ce résultat n'est pas surprenant. Lee et Nelder (2004) soulignent qu'un modèle conditionnel peut produire une moyenne conditionnelle (modèle EF) ou une moyenne marginale (modèle EA).

Le modèle EA est complété par les hypothèses suivantes : $v_i \sim \mathcal{N}(0, \sigma_v)$ et $u_{ij} \sim \mathcal{N}(0, \sigma_u)$ pour obtenir un modèle normal multivarié $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ avec une structure

de covariance spécifique, $\boldsymbol{\Sigma}$.

L'estimateur du modèle EA peut s'obtenir en maximisant la vraisemblance du modèle normal multivarié, ce qui revient à minimiser la fonction de risque suivante :

$$R_{EA}^{MCO}(\boldsymbol{\beta}, \boldsymbol{\eta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2, \tag{1.28}$$

où la norme est définie par $\|\boldsymbol{u}\|^2 = \boldsymbol{u}^\mathsf{T}\boldsymbol{u}$. Comme dans le cadre de la régression linéaire simple, la solution du problème (1.28) est obtenue par la méthode des moindres carrés ordinaires (MCO). L'estimateur MCO est défini par :

$$\widehat{\boldsymbol{\beta}}_{MCO} = \left[\sum_{i=1}^{n} \boldsymbol{X}_i^\mathsf{T}\boldsymbol{X}_i\right]^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i^\mathsf{T}\boldsymbol{y}_i.$$

L'estimateur MCO est convergent mais pas efficace (Greene, 2011; Arellano, 2003; Hsiao, 2015). Il ne tient pas compte de la structure de dépendance des données. Il a été suggéré, pour un estimateur efficace, de minimiser plutôt la fonction objectif suivante :

$$R_{EA}^{GLS}(\boldsymbol{\beta}, \boldsymbol{\eta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_{\boldsymbol{\Sigma}^{-1}}^2, \tag{1.29}$$

où la norme est définie par $\|\boldsymbol{u}\|_{\boldsymbol{\Sigma}^{-1}}^2 = \boldsymbol{u}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{u}$. L'estimateur qui en découle est l'estimateur des moindres carrés généralisés ou le «generalized least square (GLS) estimator». Il est défini par :

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left[\boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right]^{-1}\boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{y} = \left[\sum_{i=1}^{n} \boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{X}_i^*\right]^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{y}_i^*.$$

L'estimateur GLS peut se générer également par transformation. Pour ce faire, il faut multiplier les données initiales par la matrice $\boldsymbol{\Sigma}^{-1/2}$. Puis, appliquer la méthode MCO sur les données transformées, $\left\{\boldsymbol{y}^* = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}, \boldsymbol{X}^* = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}\right\}$.

Sous certaines conditions de régularité, l'estimateur GLS est sans biais, convergent, efficace et asymptotiquement gaussien (Greene, 2011). Le théorème de Gauss-Markov démontre que l'estimateur GLS est «BLUE» (Best Linear Unbiaised Estimator), c'est-à-dire qu'il est le meilleur estimateur sans biais dans la classe des estimateurs linéaires. Il est possible de dériver l'estimateur GLS en maximisant une fonction de vraisemblance (Greene, 2011). La matrice de variance-covariance de l'estimateur GLS est donnée par

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}_{GLS}) = \boldsymbol{X}^{\intercal}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}.$$

Pour estimer la matrice de variance-covariance de l'estimateur GLS, il suffit d'estimer la matrice de variance-covariance $\boldsymbol{\Sigma}$. Plusieurs estimateurs sont proposés en fonction de l'hypothèse sur le degré de dépendance et le degré d'hétérogénéité des données. Pour plus de détails, voir (Greene, 2011).

Nous terminons cette section par la présentation de l'estimateur du modèle EA proposé par Henderson (1950). L'approche d'Henderson est une procédure hybride à la croisée de l'approche fréquentiste et de l'approche bayésienne. Comme le rapporte Robinson (1991), Henderson explique que son estimateur ne devrait pas être considéré comme un estimateur de maximum de vraisemblance, même s'il provient de la maximisation d'une vraisemblance jointe.

Dans le cadre du modèle EA et sous l'hypothèse que $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{v}})$ et $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{u}})$, Henderson propose de maximiser la fonction de densité jointe de $\boldsymbol{y}$ et $\boldsymbol{v}$

$$(2\pi\sigma^2)^{-\frac{1}{2}n-\frac{1}{2}q}\begin{vmatrix}\boldsymbol{\Sigma}_{\boldsymbol{v}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}_{\boldsymbol{u}}\end{vmatrix}^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}\begin{pmatrix}\boldsymbol{v}\\ \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{v}\end{pmatrix}^{\intercal}\begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{v}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}_{\boldsymbol{u}}\end{bmatrix}^{-1}\begin{pmatrix}\boldsymbol{v}\\ \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{v}\end{pmatrix}\right\}$$

par rapport à $\boldsymbol{\beta}$ et $\boldsymbol{v}$. La maximisation de cette pseudo vraisemblance revient à minimiser la fonction objectif suivante :

$$\begin{aligned} R_{EA}(\boldsymbol{\beta}, \boldsymbol{v}) &= \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v} \end{pmatrix}^{\mathsf{T}} \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{v}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\boldsymbol{u}} \end{bmatrix}^{-1} \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v} \end{pmatrix} \\ &= \boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1}\boldsymbol{v} + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v})^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v}). \end{aligned}$$

L'estimateur des effets aléatoires gaussien dérivé de cette optimisation est BLUE et a pour expression :

$$\widehat{\boldsymbol{\beta}}_H = [\boldsymbol{X}^{\mathsf{T}}(\boldsymbol{\Sigma}_u + \boldsymbol{Z}\boldsymbol{\Sigma}_{\boldsymbol{v}}\boldsymbol{Z}^{\mathsf{T}})^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}^{\mathsf{T}}(\boldsymbol{\Sigma}_u + \boldsymbol{Z}\boldsymbol{\Sigma}_{\boldsymbol{v}}\boldsymbol{Z}^{\mathsf{T}})^{-1}\boldsymbol{y}. \tag{1.30}$$

L'estimateur $\widehat{\boldsymbol{\beta}}_H$ peut être considéré comme un estimateur bayésien ou comme un estimateur pénalisé des moindres carrés (Robinson, 1991). En effet, Koenker (2004) a montré que l'estimateur des effets aléatoires gaussien d'Henderson minimise la fonction objectif suivante :

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v}\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}} + \|\boldsymbol{v}\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1}}.$$

La reformulation de l'estimateur d'Henderson comme la solution d'une régression pénalisée suggère l'existence d'une classe plus large d'estimateurs.

La pénalité rajoutée au modèle EA permet de comprimer le paramètre individuel autour de la valeur nulle. Cette compression a l'avantage de contrôler la variabilité introduite par le paramètre d'incidence et de réduire simultanément sa dimension de nature infinie lorsque la taille de l'échantillon augmente. Il y a deux pénalités prédominantes, l_1 et l_2, qui sont couramment utilisées dans la littérature (Friedman *et al.*, 2010). Les deux pénalités sont utiles, mais leur impact est très différent dans la pratique. Lorsqu'il s'agit de sélectionner des variables et de trouver une solution

dite «sparse», la pénalité l_1 est mieux adaptée (Tibshirani, 1996). Dans le processus de sélection, le degré de compression est contrôlé par le paramètre de régularisation dont la valeur optimale est difficile à trouver.

En haute dimension, la valeur optimale du paramètre de régularisation peut être déterminée par la méthode de la validation croisée. La méthode de la validation croisée consiste à partitionner l'ensemble des données en deux classes : l'ensemble des données d'apprentissage avec lequel le modèle est ajusté et l'ensemble de données d'essai avec lequel la puissance prédictive du modèle est évaluée. Les détails de cette approche se trouvent dans l'excellent livre de (Hastie *et al.*, 2001). Cependant, l'efficacité de la méthode de la validation croisée nécessite la disponibilité de grandes quantités de données. Pour cette raison, d'autres critères tels que la validation croisée généralisée (GCV), le critère d'information d'Akaike (AIC) ou le critère d'information de Bayes (BIC) sont souvent recommandés. Le BIC est particulièrement apprécié, entre autres, pour sa consistance.

Nous présentons, ici, deux critères bayésiens pour permettre de sélectionner le meilleur paramètre de régularisation. Le premier critère est le critère de sélection de Wang et Leng (2007) connu sous le nom d'approximation des moindres carrés (AMC) et défini par :

$$\mathrm{BIC}_{1\lambda} = (\widehat{\boldsymbol{\beta}}_\lambda - \widehat{\boldsymbol{\beta}}_0)^\intercal \widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\beta}}_0)(\widehat{\boldsymbol{\beta}}_\lambda - \widehat{\boldsymbol{\beta}}_0) + (nm)^{-1}\mathrm{df}_\lambda \log(nm). \tag{1.31}$$

Les estimateurs $\widehat{\boldsymbol{\beta}}_\lambda$ et $\widehat{\boldsymbol{\beta}}_0$ sont respectivement les estimateurs avec et sans pénalité, $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}_0) = \widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}_0)$ et df_λ représente le nombre de degré de liberté.

Le second critère de sélection est un critère proposé par Wang *et al.* (2007a,b) dans

le cadre de la régression quantile. Son expression est donnée par :

$$\mathrm{BIC}_{2\lambda} = (nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} r_\tau(y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}} \widehat{\boldsymbol{\beta}}_{\tau\lambda} - \eta_i) + \frac{1}{2}(nm)^{-1} \mathrm{df}_\lambda \log(nm), \qquad (1.32)$$

où r_τ est la fonction de perte asymétrique l_1 et df_λ une mesure de la dimensionnalité effective du modèle ajusté. Dans leur article, Galvao et Montes-Rojas (2010) ont choisi comme degrés de liberté df_λ la dimension de l'ensemble $\{\boldsymbol{\beta} \cup \{i| \, |\eta_i| > \kappa\}\}$ pour les deux critères. Le paramètre κ est un nombre choisi suffisamment petit. Galvao et Montes-Rojas (2010) ont sélectionné trois valeurs différentes de $\kappa \in \{0.001, 0.01, 0.1\}$ pour évaluer la sensibilité relative des critères. Ils ont montré que les trois valeurs menaient aux mêmes conclusions.

Dans le chapitre suivant nous présentons le permier article de notre thèse qui est soumis pour publication. Dans ce premier article nous combinons la régression expectile à la méthode des estimations des équations généralisées (GEE) pour l'analyse des données longitudinales.

CHAPITRE II

PREMIER ARTICLE : WEIGHTED ASYMMETRIC LEAST SQUARES REGRESSION FOR LONGITUDINAL DATA USING GEE

Amadou Barry, Karim Oualkacha, Arthur Charpentier

Abstract.

The well-known generalized estimating equations (GEE) is widely used to estimate the effect of the covariates on the mean of the response variable. We apply the GEE method using the asymmetric least squares regression (expectile) to analyze the longitudinal data. Expectile regression naturally extends the classical least squares method and has properties similar to quantile regression. For instance, Expectile regression allows the study of the heterogeneity of the effects of the covariates over the entire distribution of the response variable, while also accounting for unobserved heterogeneity. In this paper, we present the generalized expectile estimating equations estimators, we derive their asymptotic properties and we propose a robust estimator of their variance-covariance matrix for inference. The performance of the new estimators is evaluated through exhaustive simulation studies, and their advantages in relation to existing methods are highlighted. Finally, the labor pain dataset is analyzed to illustrate the usefulness of the proposed model.

Keywords : Expectile regression, Quantile regression, GEE, Longitudinal data.

2.1 Introduction

Longitudinal and clustered data arise in many application fields such as epidemiology (Smith *et al.*, 2015), genetics (Furlotte *et al.*, 2012), economics (Hsiao, 2007), and other fields of biological and social sciences. They are characterized by the presence of a within-subject dependence that gives them desirable properties over the cross-sectional data. However, that dependency makes the statistical analysis challenging and needs to be addressed in order to generate unbiased and highly efficient estimators. Generalized estimating equations (GEE) (Liang et Zeger, 1986) is a commonly used method for the analysis of such data within a marginal model framework.
A marginal model estimates the expectation of the marginal distribution of the outcome without specifying the within-subject dependence, like cross-sectional models (Fitzmaurice *et al.*, 2008; Diggle *et al.*, 2013). The GEE model completes the marginal model by introducing a "working" covariance matrix in the estimation process to account for the within-subject dependence. As a result, the GEE yields a consistent estimator with high efficiency even with misspecification of the true covariance structure (Liang et Zeger, 1986). The GEE model estimates only the effects of the covariates on the expectation of the marginal distribution of the outcome. The expectation is a very important summary statistic, but limiting the study of the effects of the covariates to this is not enough unless the covariates uniformly affects the whole distribution of the response variable. With its favorable properties, the GEE can be extended beyond the mean using the expectile regression (ER).

ER models the relationship between the covariates and the response variable by estimating the effect of the predictors at different points of the conditional cumulative distribution function (c.d.f.) of the response variable. These points are generally

the mean (expectile of level 0.5), and other expectiles above and below the mean. ER estimates the impact of the covariates on the location, scale, and shape of the response distribution. By doing so, the ER captures the heterogeneity of the effects of the covariates on the response variable. For example, when covariates affect the mean and the tail of the distribution in different ways.

ER was introduced in 1976 by Aigner *et al.* (1976) under a likelihood-based approach. A decade later, Newey et Powell (1987) presented a detailed study of this new class of estimator. They presented its favorable properties, like its location and scale equivariance property, and derived its asymptotic properties. After the Efron (1991) paper, expectile lived quietly in the shadows for decades, as mentioned in (Eilers, 2013). Recently, it came into the spotlight. After its re-discovery, early contributions to ER focused on the application of the ER method to spline and smoothing model (Schnabel et Eilers, 2009; Rossi et Harvey, 2009; Sobotka et Kneib, 2012; Sobotka *et al.*, 2013). Other works focused on contrasting ER and quantile regression (QR), on showing how to get quantiles from a fine grid of expectiles, on the crossing curves problem and on promoting application of ER (Kneib, 2013a; Schnabel et Eilers, 2013; Waltrup *et al.*, 2015a). Today ER is extended to many classes of models, such as Bayesian (Majumdar et Paul, 2016; Waldmann *et al.*, 2016; Xing et Qian, 2017), nonparametric (Righi *et al.*, 2014; Yang et Zou, 2015), nonlinear (Kim et Lee, 2016), neural network (Xu *et al.*, 2016; Jiang *et al.*, 2017), and support vector machine (Farooq et Steinwart, 2017). Recently, Schulze et Kauermann (2015) combined smoothing and random intercept to fit clustered data with penalized splines.

ER generalizes mean regression in the same way that quantile regression (QR) (Koenker et Bassett, 1978) generalizes median regression. The QR method was adapted to longitudinal data using GEE approach. The main idea consists of smoothing the QR

estimating functions in order to make them differentiable with respect to regression parameters. Jung (1996) proposed the quasi-likelihood approach to analyze the median regression model for dependent data. Chen *et al.* (2004) derived a QR estimator for correlated data using GEE approach based on independence working correlation. Along the same lines, Fu et Wang (2012) combined the between- and within-subject estimating functions to account for the correlations between repeated measurements in the estimation of QR model. Lu et Fan (2015) proposed a general stationary autocorrelation matrix for the working correlation to enhance the efficiency of the QR inference.

Both QR and ER provide an overview of the effects of the covariates on the distribution of the response variable. Their resemblance and usefulness have already been discussed in the literature, see for examples (Efron, 1991; Kneib, 2013c; Waltrup *et al.*, 2015a).

This paper makes its contribution by introducing a new class of estimators for the analysis of dependent data. Section 2.2 defines the expectile statistic, and introduces the ER method for cross-sectional data and the generalized expectile estimating equation (GEEE) method for longitudinal data. In Section 2.3, the asymptotic properties of the GEEE estimator of the model parameters and an estimator of its variance-covariance matrix are presented. The evaluation of the small sample performance of the estimators, through extensive simulation studies, is presented in Section 2.4. The GEEE estimator is applied to a real data set and the results are presented in Section 2.5 ; the conclusion is presented in Section 2.6. All proofs are available in the Appendix.

2.2 Models and Methods

This Section introduces the univariate expectile and the ER model.

2.2.1 Expectile and expectile regression

Expectile of a random variable Y is defined as the solution $\mu_\tau(Y)$ which minimizes the loss function

$$\mathbb{E}\{\rho_\tau(Y-\theta)\} \tag{2.1}$$

over $\theta \in \mathbb{R}$ for a fixed value of $\tau \in (0,1)$. The function $\rho_\tau(\cdot)$, of the form

$$\rho_\tau(t) = |\tau - \mathbb{1}(t \leq 0)| \cdot t^2$$

is the asymmetric square loss function that assigns weights τ and $1-\tau$ to positive and negative deviations, respectively.

By equating the first derivative of (2.1) to zero, the expectile can also be defined as solution of

$$\mu_\tau(Y) = \mu_\tau = \mu - \frac{1-2\tau}{1-\tau}\,\mathbb{E}\left[\{Y-\mu_\tau(Y)\}\mathbb{1}\{Y > \mu_\tau(Y)\}\right], \tag{2.2}$$

where $\mu = \mu_{0.5}(Y) = \mathbb{E}(Y)$. This definition, presented by Newey et Powell (1987), shows that expectile is determined by the tail expectations of the distribution of Y. Interestingly, we found that expectile can be defined as

$$\mu_\tau = \mathbb{E}\left[\frac{\psi_\tau(Y-\mu_\tau)}{\mathbb{E}\left[\psi_\tau(Y-\mu_\tau)\right]}Y\right],$$

where $\psi_\tau(t) = |\tau - \mathbb{1}(t \leq 0)|$ is the check function. This latter definition, which is much more meaningful in the context of regression, reveals that expectiles, like the

mean, are weighted averages.

Given a random sample, $\{(y_i)\}_{i=1}^n$, the τ-th empirical expectile

$$\widehat{\mu}_\tau = \sum_{i=1}^{n} \frac{\psi_\tau(y_i - \widehat{\mu}_\tau)}{\sum_{i=1}^n \psi_\tau(y_i - \widehat{\mu}_\tau)} y_i$$

is the solution which minimizes the empirical loss function

$$\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_i - \theta). \tag{2.3}$$

Newey et Powell (1987) have shown that expectile function has attractive properties. Expectile summarizes the c.d.f. of Y in the same way that quantile does. Moreover, expectile is location and scale equivariant, that is for $s > 0$ and $t \in \mathbb{R}$, $\mu_\tau(sY + t) = s\mu_\tau(Y) + t$. More details about the properties of expectile and results on ER are given by Efron (1991).

To introduce the ER method, consider the classical linear regression

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \tag{2.4}$$

where y_i is the scalar response, ε_i is the random error, $\boldsymbol{x}_i \in \mathbb{R}^p$ is the vector of covariates and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the unknown parameter that needs to be estimated. Under this framework, Newey et Powell (1987) introduced the ER model for a fixed $\tau \in (0, 1)$ as

$$\mu_\tau(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau, \quad \text{with} \quad \mu_\tau(\varepsilon_i) = 0. \tag{2.5}$$

The assumption $\mu_\tau(\varepsilon_i) = 0$ ensures that the random error is centered on the τ-th expectile. The corresponding ER estimator is defined as the unique solution which minimizes the objective function

$$\frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau) \tag{2.6}$$

over $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$. The asymmetric loss function associated with the expectile function is continuously differentiable, and solving equation (2.6) gives

$$\widehat{\boldsymbol{\beta}}_\tau = \Big(\sum_{i=1}^{n} \boldsymbol{x}_i{}^{\mathsf{T}} \psi_\tau(\widehat{\varepsilon}_{it\tau}) \boldsymbol{x}_i \Big)^{-1} \Big(\sum_{i=1}^{n} \boldsymbol{x}_i \psi_\tau(\widehat{\varepsilon}_{it\tau}) y_i \Big), \tag{2.7}$$

where $\widehat{\varepsilon}_{it\tau} = y_i - \boldsymbol{x}_i{}^{\mathsf{T}} \widehat{\boldsymbol{\beta}}_\tau$. The ER estimator is easily computed with the iterated weighted least squares algorithm. In addition to deriving the asymptotic properties of the above ER estimator, Newey et Powell (1987) proposed a robust estimator of the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}_\tau$.

Note that the ER estimator was presented previously by Aigner *et al.* (1976) through a likelihood-based approach. The likelihood is derived by assuming an asymmetric normal distribution (AND) for the disturbance, $u \sim \text{AND}(u; \mu_\tau, \sigma^2, \tau)$, with density

$$f(u; \mu_\tau, \sigma^2, \tau) = \frac{2}{\sqrt{\pi\sigma^2}} \frac{\sqrt{\tau(1-\tau)}}{\sqrt{\tau} + \sqrt{1-\tau}} \exp\left\{ -\rho_\tau\left(\frac{u - \mu_\tau}{\sigma} \right) \right\}, \tag{2.8}$$

where $\mu_\tau, \sigma,$ and τ are respectively the location, scale and asymmetric parameters. Now, substitute $\mu_{i\tau} = \boldsymbol{x}_i{}^{\mathsf{T}} \boldsymbol{\beta}_\tau$ and assume the n observations are independents, then the ER estimator is equivalent to the maximum of the likelihood function

$$\mathbb{L}(\boldsymbol{\beta}; \sigma, \tau, \boldsymbol{y}) \propto \sigma^{-2n} \exp\left\{ -\sum_{i=1}^{n} \rho_\tau\left(\frac{y_i - \boldsymbol{x}_i{}^{\mathsf{T}} \boldsymbol{\beta}_\tau}{\sigma} \right) \right\},$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^{\mathsf{T}}$. The AND distribution is not to be confused with the class of density functions related to the standard density function and proposed by Azzalini (1985).

2.2.2 GEEE for longitudinal data

This Section presents the model and method of the GEEE for longitudinal data. Consider the data $\{y_{it}, \boldsymbol{x}_{it}\}_{1\leq i\leq n, 1\leq t\leq m_i}$ generated by the following model

$$y_{it} = \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta} + \varepsilon_{it}, \tag{2.9}$$

where y_{it} is the t-th observation of the response variable for the i-th individual, $\boldsymbol{x}_{it} = (x_{it}^1, \ldots, x_{it}^p)$ is the $p \times 1$ covariates, ε_{it} the random error and $\boldsymbol{\beta}$ the $p \times 1$ true parameter vector that needs to be estimated. Equation model (2.9) can be conveniently represented in individual notation as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \tag{2.10}$$

where $\boldsymbol{y}_i$ is the dependent observations of the response variable of the individual i, $\boldsymbol{X}_i$ the corresponding $m_i \times p$ matrix covariates and $\boldsymbol{\varepsilon}_i$ the vector error. Individual observations can also be stacked and presented in matrix form as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.11}$$

where $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ are $N \times 1$ vectors, $\boldsymbol{X}$ is $N \times p$ matrix and $N = \sum_{i=1}^{n} m_i$.

Using the location-scale equivariance property of the expectile function, the corresponding conditional expectile of level τ of the model equation (2.9) is defined as

$$\mu_\tau(y_{it}|\boldsymbol{x}_{it}) = \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}_\tau, \qquad \mu_\tau(\varepsilon_{it}) = 0. \tag{2.12}$$

The assumption, $\mu_\tau(\varepsilon_{it}) = 0$, is introduced to guarantee that the random error is centered on the τ-th expectile. The parameter $\boldsymbol{\beta}_\tau$ measures the effect of the covariates $\boldsymbol{x}_{it}$ on the location, scale and shape of the conditional distribution of the response.

A practical estimator of the parameter can be obtained by looking for the solution of the following expectile estimating equations

$$\boldsymbol{S}_I(\boldsymbol{\beta}_\tau) = \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_\tau)\Big[\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_\tau\Big] = \mathbf{0}, \tag{2.13}$$

where $\boldsymbol{\Psi}_\tau(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_\tau) = \text{diag}\Big(\psi_\tau(y_{i1} - \boldsymbol{x}_{i1}{}^{\mathsf{T}}\boldsymbol{\beta}_\tau), \ldots, \psi_\tau(y_{im_i} - \boldsymbol{x}_{im_i}{}^{\mathsf{T}}\boldsymbol{\beta}_\tau)\Big)$. The resulting estimator $\widehat{\boldsymbol{\beta}}_{I\tau}$ can also be derived as the minimizer of the following objective function

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{m_i} \rho_\tau\Big(y_{it} - \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}_\tau\Big) \tag{2.14}$$

over $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$. The explicit form of the resulting estimator $\widehat{\boldsymbol{\beta}}_{I\tau}$ is similar to (2.7).

When $\tau = 0.5$, the estimator $\widehat{\boldsymbol{\beta}}_{I\tau}$ corresponds to the GEE estimator introduced by Liang et Zeger (1986) with an independent working correlation between observations from the same subject. This fact is exploited to extend the GEE to the generalized expectile estimating equation (GEEE).

The GEEE method models the underlying correlation structure from the same subject by formally including a hypothesized structure with the within-subject correlation. For a fixed τ, the GEEE estimator $\widehat{\boldsymbol{\beta}}_\tau$ is derived by solving the following GEEE equations

$$\boldsymbol{S}(\boldsymbol{\beta}_\tau) = \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{V}_{i\tau}^{-1} \boldsymbol{\Psi}_\tau(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_\tau)\Big[\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_\tau\Big] = \mathbf{0}, \tag{2.15}$$

where $\boldsymbol{V}_{i\tau}$ is a working covariance matrix represented as

$$\boldsymbol{V}_{i\tau} = \sigma_\tau^2 \boldsymbol{A}_{i\tau}^{\frac{1}{2}} \boldsymbol{R}_i(\boldsymbol{\alpha}_\tau) \boldsymbol{A}_{i\tau}^{\frac{1}{2}}, \tag{2.16}$$

and σ_τ^2 is the nuisance parameter. $\boldsymbol{A}_{i\tau}$ is the $m_i \times m_i$ diagonal matrix with the variance function $\nu(\mu_i)$ as diagonal elements and $\boldsymbol{R}_i(\boldsymbol{\alpha}_\tau)$ as the working correlation matrix.

The working correlation matrix $\boldsymbol{R}_i(\boldsymbol{\alpha}_\tau)$ describes the correlation pattern of within-subject observations with the $K \times 1$ vector parameter $\boldsymbol{\alpha}_\tau$. Liang et Zeger (1986) proposed several types of working correlation structures (independent, exchangeable, autoregressive, unstructured, etc.) for the case $\tau = 0.5$. These working correlations are adapted and extended to the GEEE approach. The extension of some of the most common and popular ones are presented below.

The GEEE independent working correlation structure is the simplest form of working correlation with identity matrix and is the structure assumed by the expectile estimating equations model (2.13). The GEEE exchangeable structure is a simple extension of the independence working correlation. It assumes a common correlation, $\rho_{ts\tau} = \alpha_\tau, \forall t \neq s$, between any pair of measurements. The GEEE AR1 structure correlation defines the correlation of a pair of observations as a decreasing function of their distance in time, $\rho_{ts\tau} = \alpha_\tau^{|t-s|}$. This structure assigns the highest correlation to adjacent pairs of observation and the lowest correlation to distant pairs. The GEEE unstructured, as its name suggests, imposes no structure to the correlation matrix and defines the correlations of all pairs of measurements differently, without any explicit pattern, $\rho_{ts\tau} = \alpha_{ts\tau}, \forall t \neq s$.

All these types of working correlation are usually unknown and must be estimated. They are estimated in the iterative fitting process using the current value of the parameter vector. Indeed, the estimators can be computed as iterated weighted least

squares estimators. The estimation algorithm for the GEEE exchangeable working correlation can be summarized as the following stepwise procedure.

Algorithm : The GEEE algorithm

Step 1. Let $\widetilde{\boldsymbol{\beta}}_\tau^{(0)} = \widehat{\boldsymbol{\beta}}_{I\tau}$, the estimator obtained from (2.13).

Step 2. Given $\widetilde{\boldsymbol{\beta}}_\tau^{(r-1)}$ from the $r-1$ step update

$$\widehat{\sigma}_\tau^{2(r)} \leftarrow \frac{1}{N-p}\sum_{i=1}^{n}\sum_{t=1}^{m_i}\psi_\tau(\widehat{\varepsilon}_{it\tau})^2\widehat{\varepsilon}_{it\tau}^2,$$

$$\widehat{\alpha}_\tau^{(r)} \leftarrow \frac{1}{(N_1-p)\widehat{\sigma}_\tau^{2(r)}}\sum_{i=1}^{n}\sum_{t<s}^{m_i}\psi_\tau(\widehat{\varepsilon}_{it\tau})\widehat{\varepsilon}_{it\tau}\psi_\tau(\widehat{\varepsilon}_{is\tau})\widehat{\varepsilon}_{is\tau},$$

where $N_1 = \frac{1}{2}\sum_{i=1}^{n} m_i(m_i-1)$ and $\widehat{\varepsilon}_{it\tau} = y_{it} - \boldsymbol{x}_{it}{}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_\tau^{(r-1)}$.

Step 3. Update $\widehat{\boldsymbol{\beta}}_\tau^{(r)}$ by

$$\widehat{\boldsymbol{\beta}}_\tau^{(r)} \leftarrow \widehat{\boldsymbol{\beta}}_\tau^{(r-1)} + \Big[\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{V}_{i\tau}^{-1}(\widehat{\alpha}_\tau^{(r-1)})\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\beta}}_\tau^{(r-1)})\boldsymbol{X}_i\Big]^{-1}\boldsymbol{S}(\widehat{\boldsymbol{\alpha}}_\tau^{(r-1)}, \widehat{\boldsymbol{\beta}}_\tau^{(r-1)}),$$

where $\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\beta}}_\tau^{(r-1)}) = \boldsymbol{\Psi}_\tau(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_\tau^{(r-1)})$.

Step 4. Repeat the above iteration, **Steps** 2-3, until convergence.

The algorithm also applies to other types of working correlation; simply choose the appropriate estimator of the parameter $\boldsymbol{\alpha}$ which is either a scalar or a vector, depending on the type of correlation. For example, for a GEEE autoregressive AR1 working correlation structure, the scalar parameter α_τ is estimated by

$$\widehat{\alpha}_\tau = \frac{1}{(N_2-p)\sigma_\tau^2}\sum_{i=1}^{n}\sum_{t<m_i-1}\psi_\tau(\widehat{\varepsilon}_{it\tau})\widehat{\varepsilon}_{it\tau}\psi_\tau(\widehat{\varepsilon}_{i,t+1,\tau})\widehat{\varepsilon}_{i,t+1,\tau}, \quad N_2 = \sum_{i=1}^{n}(m_i-1).$$

For a GEEE unstructured working correlation structure, every element of the $m_i(m_i+1)/2$-vector parameter $\boldsymbol{\alpha}_\tau$ is estimated by

$$\widehat{\alpha}_{ts\tau} = \frac{1}{(N-p)\sigma_\tau^2} \sum_{i=1}^{n} \psi_\tau(\widehat{\varepsilon}_{it\tau})\widehat{\varepsilon}_{it\tau}\psi_\tau(\widehat{\varepsilon}_{is\tau})\widehat{\varepsilon}_{is\tau}.$$

Generalization to other GEEE-working correlations is straightforward. The algorithm converges rapidly, usually after 5 iterations, for all the correlation structures, except for the Un working correlation. Similar to the GEE algorithm ($\tau = 0.5$), (Hardin et Hilbe, 2012, chap. 2, p.78) the algorithm of the unrestricted working correlation structure is not always guaranteed to converge, especially when the number of within-subject observations increase or when the dataset is unbalanced.

In Section 2.3, it is shown that the GEEE estimator $\widehat{\boldsymbol{\beta}}_\tau$ is consistent and asymptotically normally distributed. In addition, the simulation results of Section 2.4 show that the GEEE method yields a consistent and highly efficient estimator even with a misspecification of the true covariance structure.

GEEE for a sequence of expectiles

The sequence of expectiles is often necessary, usually the mean and a few expectiles above and below the mean, to describe the effect of the covariates on the conditional distribution of the response variable. Also, the simultaneous estimation allows them to share strength among each other and to gain better estimation accuracy than individually estimated (Liu et Wu, 2011). For a fixed $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)$ the GEEE

estimating functions are defined as

$$\begin{aligned}\boldsymbol{S}(\boldsymbol{\beta}_{\boldsymbol{\tau}}) &= \sum_{k=1}^{q} \boldsymbol{S}_{\tau_k}(\boldsymbol{\beta}_{\tau_k}) \\ &= \sum_{i=1}^{n} (\boldsymbol{W} \otimes \boldsymbol{X}_i)^{\top} \boldsymbol{V}_{i\boldsymbol{\tau}}^{-1} \boldsymbol{\Psi}_{\boldsymbol{\tau}} \Big(\mathbb{1}_q \otimes \boldsymbol{y}_i - (\mathbb{I}_q \otimes \boldsymbol{X}_i)\boldsymbol{\beta}_{\boldsymbol{\tau}} \Big) \Big[\mathbb{1}_q \otimes \boldsymbol{y}_i - (\mathbb{I}_q \otimes \boldsymbol{X}_i)\boldsymbol{\beta}_{\boldsymbol{\tau}} \Big],\end{aligned} \tag{2.17}$$

where $\boldsymbol{S}_{\tau_k}$ is defined in (2.15), and $\boldsymbol{W} = [\text{diag}(w_k)]_{k=1}^{q}$ is the $q \times q$ matrix of weights controlling the relative influence of the q expectiles. $\boldsymbol{V}_{i\boldsymbol{\tau}} = [\text{diag}(\boldsymbol{V}_{i\tau_k})]_{k=1}^{q}$ is a $qm_i \times qm_i$ block-diagonal working covariance matrix. For any fixed τ_k, the expression of the $m_i \times m_i$ matrix $\boldsymbol{V}_{i\tau_k}$ is given by (2.16) and

$$\boldsymbol{\Psi}_{\boldsymbol{\tau}} \Big(\mathbb{1}_q \otimes \boldsymbol{y}_i - (\mathbb{I}_q \otimes \boldsymbol{X}_i)\boldsymbol{\beta}_{\boldsymbol{\tau}} \Big) = \text{diag} \Big(\boldsymbol{\Psi}_{\tau_1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_{\tau_1}), \ldots, \boldsymbol{\Psi}_{\tau_q}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_{\tau_q}) \Big).$$

The parameter $\boldsymbol{\beta}_{\boldsymbol{\tau}}$ is obtained using the iterative re-weighted least squares algorithm, as shown above, for a single expectile. We used equation (2.17) to derive the asymptotic properties of the GEEE estimators and to estimate the variance-covariance matrix of the GEEE estimators in the real data application. However, we obtain the same GEEE parameter estimates using equation (2.15) or equation (2.17).

In the next Section, the asymptotic properties of the GEEE estimator are presented for a sequence of expectiles.

2.3 Asymptotic properties

This Section presents the asymptotic properties of the GEEE estimator for several fixed expectiles $\boldsymbol{\tau}$. In the first step, the asymptotic properties of the GEEE estimator $\widehat{\boldsymbol{\beta}}_{I\boldsymbol{\tau}}$ with the independent working correlation structure are presented. Subsequently,

the asymptotic properties of the GEEE estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}}$ with a general correlation structure are derived. The main reason for presenting the results of $\widehat{\boldsymbol{\beta}}_{I\boldsymbol{\tau}}$ separately is that ; it is also the estimator of the expectile regression for a marginal model based on the AND distribution (2.8). In the following Section, we assume that $n \to \infty$ and that $m = \max_{1\leq i\leq n} m_i$ is fixed. The proof of all results can be found in the Appendix.

2.3.1 Asymptotic properties for the independent GEEE

To begin, assume the following conditions.

A1. The data $\{(\boldsymbol{y}_i, \boldsymbol{X}_i)\}_{i=1}^n$ are independent across i, and

$\text{Var}\left[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}})\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}\right] = \mathbb{E}\left[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}})\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}{}^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}})\right] = \boldsymbol{\Sigma}_{i\boldsymbol{\tau}}$, where $\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}} = \left(\boldsymbol{\varepsilon}_{i\tau_1}{}^{\mathsf{T}}, \ldots, \boldsymbol{\varepsilon}_{i\tau_q}{}^{\mathsf{T}}\right)^{\mathsf{T}}$

$\boldsymbol{\varepsilon}_{i\tau_k} = (\varepsilon_{i1\tau_k}, \ldots, \varepsilon_{im_i\tau_k})^{\mathsf{T}}$, $\varepsilon_{it\tau_k} = y_{it} - \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}_{\tau_k}$ and $\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}) = \left[\text{diag}(\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{i\tau_k}))\right]_{k=1}^{q}$.

A2. The limiting forms of the following matrices are positive definite

$$\boldsymbol{D}_{I1}(\boldsymbol{\tau}) = \lim_{n\to\infty} N^{-1}\sum_{i=1}^{n}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^{\mathsf{T}}\,\mathbb{E}[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}})](\mathbf{I}_q\otimes\boldsymbol{X}_i),$$

$$\boldsymbol{D}_{I0}(\boldsymbol{\tau}) = \lim_{n\to\infty} N^{-1}\sum_{i=1}^{n}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^{\mathsf{T}}\boldsymbol{\Sigma}_{i\boldsymbol{\tau}}(\boldsymbol{W}\otimes\boldsymbol{X}_i).$$

A3. $\max_{\substack{1\leq i\leq n\\ 1\leq t\leq m_i}} \|x_{it}\| < M$.

Assumptions **A1-A3** are standard for longitudinal data models. Condition **A1** ensures independence across individuals, but permits a within-dependency between observations of the same individual and allows heterogeneity across individuals.

Condition **A2** is a standard full rank condition. Observe that, when $\tau = 1/2$, then $\boldsymbol{\Sigma}_{i0.5} = 1/4\,\mathrm{Var}[\boldsymbol{\varepsilon}_{i0.5}]$ becomes the variance of $\boldsymbol{\varepsilon}_i$ up to a factor and $\boldsymbol{D}_{I0} = 1/4 \lim_{n\to\infty} N^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\top} \mathrm{Var}[\boldsymbol{\varepsilon}_{i0.5}] \boldsymbol{X}_i$. Considering $\boldsymbol{D}_{I1} = 1/2 \lim_{n\to\infty} N^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\top} \boldsymbol{X}_i$, we see that this factor disappears in the expression of the variance of the estimator. Therefore, when $\tau = 1/2$, the condition **A2** is reduced to a condition on the matrices $N^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\top} \mathrm{Var}[\boldsymbol{\varepsilon}_i] \boldsymbol{X}_i$ and $N^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\top} \boldsymbol{X}_i$. Condition **A3** is important both for the convergence and for the Lindeberg condition. The following **Theorem** states the results of the asymptotic properties of the GEEE estimator $\widehat{\boldsymbol{\beta}}_{I\boldsymbol{\tau}}$ assuming an independent working correlation structure.

Theorem 2.3.1. *Assume that $\widehat{\boldsymbol{\beta}}_{I\boldsymbol{\tau}}$ is the solution of the estimating function (2.13) and suppose the data are generated by model (2.9), and that conditions **A1-A3** are satisfied. If $\mathbb{E}|\psi_{\tau_k}(\varepsilon_{it\tau_k})|^{4+\nu} < \Delta$ and $\mathbb{E}|\varepsilon_{it\tau_k}|^{4+\nu} < \Delta$ for some $\nu > 0$ and $\Delta > 0$, then for every fixed sequence of expectiles $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)$*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{I\boldsymbol{\tau}} - \boldsymbol{\beta}_{\boldsymbol{\tau}}) \xrightarrow{d} \mathcal{N}\Big(\mathbf{0}, \boldsymbol{D}_{I1}^{-1}(\boldsymbol{\tau}) \boldsymbol{D}_{I0}(\boldsymbol{\tau}) \boldsymbol{D}_{I1}^{-1}(\boldsymbol{\tau})\Big).$$

In order to use this new estimator $\widehat{\boldsymbol{\beta}}_{I\boldsymbol{\tau}}$ to make inference, an estimator of its VC-matrix is presented in **Theorem 2.3.2**. This will make it possible to construct large sample confidence intervals or hypotheses tests. This estimator is a generalization of the robust VC estimator proposed by White (1980) and used in, among other things, multilevel analysis (Liang et Zeger, 1986). This estimator inherits the same property namely in that it takes into account the within-subject-correlation and the heteroscedasticity between subjects. In sum, the proposed VC-matrix estimator is a

commonly advocated covariance matrix estimator for longitudinal data. Let,

$$\widehat{\boldsymbol{D}}_{I1}(\boldsymbol{\tau}) = N^{-1}\sum_{i=1}^{n}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}})(\mathbf{I}_q\otimes\boldsymbol{X}_i),$$

$$\widehat{\boldsymbol{D}}_{I0}(\boldsymbol{\tau}) = N^{-1}\sum_{i=1}^{n}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}}(\boldsymbol{W}\otimes\boldsymbol{X}_i)$$

where $\widehat{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} = \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}})\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}}\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}}{}^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}})$ and $\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}}$ is obtained by replacing $\boldsymbol{\beta}_{\boldsymbol{\tau}}$ with $\widehat{\boldsymbol{\beta}}_{I\boldsymbol{\tau}}$ in the expression of $\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}$. Then, we have **Theorem 2.3.2**.

Theorem 2.3.2. *Suppose the data are generated by model (2.9) and that conditions **A1-A3** are satisfied. If* $\mathbb{E}|\psi_{\tau_k}(\widehat{\varepsilon}_{it\tau_k})|^{4+\nu} < \Delta$ *and* $\mathbb{E}|\varepsilon_{it\tau_k}|^{4+\nu} < \Delta$ *for some* $\nu > 0$ *and* $\Delta > 0$, *then for every fixed sequence of expectiles* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)$

$$\widehat{\boldsymbol{D}}_{I1}^{-1}(\boldsymbol{\tau})\widehat{\boldsymbol{D}}_{I0}(\boldsymbol{\tau})\widehat{\boldsymbol{D}}_{I1}^{-1}(\boldsymbol{\tau}) \xrightarrow{p} \boldsymbol{D}_{I1}^{-1}(\boldsymbol{\tau})\boldsymbol{D}_{I0}(\boldsymbol{\tau})\boldsymbol{D}_{I1}^{-1}(\boldsymbol{\tau}).$$

2.3.2 Asymptotic properties for the general GEEE estimator

After presenting the asymptotic properties of the GEEE-independent working correlation estimator, this subsection presents the asymptotic properties of the GEEE-estimator for a general working correlation. Assume that

B1. The data $\{(\boldsymbol{y}_i, \boldsymbol{X}_i)\}_{i=1}^{n}$ are independent across i and Var $\left[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}})\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}\right] = \boldsymbol{\Sigma}_{i\boldsymbol{\tau}}$.

B2. The limiting forms of the following matrices are positive definite

$$\boldsymbol{D}_1(\boldsymbol{\tau}) = \lim_{n\to\infty} N^{-1}\sum_{i=1}^{n}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^{\mathsf{T}}\boldsymbol{V}_{i\boldsymbol{\tau}}^{-1}\,\mathbb{E}[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}})](\mathbf{I}_q\otimes\boldsymbol{X}_i),$$

$$\boldsymbol{D}_0(\boldsymbol{\tau}) = \lim_{n\to\infty} N^{-1}\sum_{i=1}^{n}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^{\mathsf{T}}\boldsymbol{V}_{i\boldsymbol{\tau}}^{-1}\boldsymbol{\Sigma}_{i\boldsymbol{\tau}}\boldsymbol{V}_{i\boldsymbol{\tau}}^{-1}(\boldsymbol{W}\otimes\boldsymbol{X}_i).$$

B3. $\max_{\substack{1\le i\le n\\ 1\le t\le m_i}} \|x_{it}\| < M.$

The following Theorem derives the asymptotic properties of the GEEE estimator with a general working correlation, under the above conditions.

Theorem 2.3.3. *Suppose the data are generated by model (2.9) and that conditions* ***B1-B3*** *are satisfied. If* $\mathbb{E}|\psi_{\tau_k}(\varepsilon_{it\tau_k})|^{4+\nu} < \Delta$ *and* $\mathbb{E}|\varepsilon_{it\tau_k}|^{4+\nu} < \Delta$ *for some* $\nu > 0$ *and* $\Delta > 0$, *then for every fixed sequence of expectiles* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)$

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} - \boldsymbol{\beta}_{\boldsymbol{\tau}}) \xrightarrow{d} \mathcal{N}\Big(\mathbf{0}, \boldsymbol{D}_1^{-1}(\boldsymbol{\tau})\boldsymbol{D}_0(\boldsymbol{\tau})\boldsymbol{D}_1^{-1}(\boldsymbol{\tau})\Big).$$

In the same way as with the GEEE-independent working correlation estimator, the following **Theorem 2.3.4** proposes an estimator of the VC-matrix of estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}}$. Consider $\widehat{\boldsymbol{V}}_{i\boldsymbol{\tau}}$ to be a consistent estimator of $\boldsymbol{V}_{i\boldsymbol{\tau}}$. Then, under the above conditions, Theorem 2.3.4 is stated as follows

Theorem 2.3.4. *Suppose the data are generated by model (2.9) and that conditions* ***B1-B3*** *are satisfied. Assume* $\mathbb{E}|\psi_{\tau_k}(\widehat{\varepsilon}_{it\tau_k})|^{4+\nu} < \Delta$ *and* $\mathbb{E}|\varepsilon_{it\tau_k}|^{4+\nu} < \Delta$ *for some* $\nu > 0$ *and* $\Delta > 0$. *Then for every fixed sequence of expectiles* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)$

$$\widehat{\boldsymbol{D}}_1^{-1}(\boldsymbol{\tau})\widehat{\boldsymbol{D}}_0(\boldsymbol{\tau})\widehat{\boldsymbol{D}}_1^{-1}(\boldsymbol{\tau}) \xrightarrow{p} \boldsymbol{D}_1^{-1}(\boldsymbol{\tau})\boldsymbol{D}_0(\boldsymbol{\tau})\boldsymbol{D}_1^{-1}(\boldsymbol{\tau})$$

where

$$\widehat{\boldsymbol{D}}_1(\boldsymbol{\tau}) = N^{-1}\sum_{i=1}^{n}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^T\widehat{\boldsymbol{V}}_{i\boldsymbol{\tau}}^{-1}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}})(\mathbf{I}_q\otimes\boldsymbol{X}_i),$$

$$\widehat{\boldsymbol{D}}_0(\boldsymbol{\tau}) = N^{-1}\sum_{i=1}^{n}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^T\widehat{\boldsymbol{V}}_{i\boldsymbol{\tau}}^{-1}\widehat{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}}\widehat{\boldsymbol{V}}_{i\boldsymbol{\tau}}^{-1}(\boldsymbol{W}\otimes\boldsymbol{X}_i),$$

and $\widehat{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} = \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}})\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}}\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}}{}^T\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}}_{i\boldsymbol{\tau}})$.

2.4 Simulations

In this Section, the small sample performance of the estimators is evaluated through extensive simulation studies. The random samples are generated from the following linear model (M_γ) :

$$y_{it} = \beta_0 + x_{it}\beta_1 + (1+\gamma x_{it})\varepsilon_{it},\ i = 1,\ldots,n \text{ and } t = 1,\ldots,m_i. \tag{2.18}$$

Two versions of model (2.18) are considered with respect to the parameter $\gamma \in \{0, 1/10\}$: a location-shift model (M_0) corresponding to $\gamma = 0$, which helps assess the performance of the estimators for an homoscedastic scenario, and a location-scale-shift model $(M_{1/10})$ corresponding to $\gamma = 1/10$, serving to assess the performance of the estimators in the presence of heteroscedasticity.

The corresponding GEEE model of (M_0) is $\mu_\tau(y_{it}) = \beta_{0\tau} + x_{it}\beta_1$ where $\beta_{0\tau} = \beta_0 + \mu_\tau(\varepsilon_{it})$, so that only the intercept term varies with τ and the expectiles functions are parallel lines. The GEEE model related to the $(M_{1/10})$ model is $\mu_\tau(y_{it}) = \beta_{0\tau} + x_{it}\beta_{1\tau}$ where $\beta_{0\tau} = \beta_0 + \mu_\tau(\varepsilon_{it})$ and $\beta_{1\tau} = \beta_1 + \gamma\mu_\tau(\varepsilon_{it})$. Therefore, in the presence of heteroscedasticity both the intercept and the slope vary with τ.

We generate the regressor x_{it} from a Gaussian distribution in (M_0) and from a Chi-square distribution in $(M_{1/10})$ and set the parameters β_0 and β_1 to 0. In order to allow for simulation of dependent errors with different marginal distributions, we first simulate dependent uniform margins from a Gaussian copula with an AR1 correlation structure. We then generate the dependent random errors as quantiles of the uniform margins from three distinct marginal distributions : Normal, Student with three degrees of freedom and Chi-square with three degrees of freedom. We also centered the random errors on the τ-th expectile. Specifically, we generate the data as follows

1. Generate x_{it} from a Gaussian distribution in (M_0) and from a Chi-square in $(M_{1/10})$;
2. Generate a uniform sample : $(u_1, \ldots, u_{m_i})$ from a Gaussian Copula with AR1 correlation structure;
3. For $t = 1, \ldots, m_i$, generate the dependent random error $\varepsilon'_{it} = F^{-1}(u_{it})$, where $F(.)$ is one of the three marginal distributions : Gaussian, Student or Chi-square distribution;
4. Center the random error : $\varepsilon_{it} = \varepsilon'_{it} - \mu_\tau(\varepsilon'_{it})$;
5. Generate the final sample : $y_{it} = \beta_0 + x_{it}\beta_1 + (1 + \gamma x_{it})\varepsilon_{it}$.

We used three different values for the AR1 correlation structure : low $\rho = 0.1$, medium $\rho = 0.5$, and high $\rho = 0.9$ correlations. Each model is produced according to two different sample sizes $n \in \{50, 100\}$. Finally, for the number of repeated measurements m_i, a balanced design with $m_i = 4$ and an unbalanced design are studied. In the unbalanced design, m_i is an number randomly generated between 3 and 8 with

equal probability. The extensive simulation is carried out with 400 replications for each parameter-combination scenario. In each scenario, the focus is on the effect of the regressor, x_{it}, at the expectiles $\tau \in \{0.25, 0.5, 0.75\}$. All computations are implemented in `R` code language (R Core Team, 2018).

The results of the GEEE regression are analyzed using four different and popular working correlation matrices : independence, exchangeable, AR1 and unstructured correlation. The average bias (Bias) and relative efficiency (EFF) in relation to the independent working correlation of the estimates are reported for the measurement of the quality of the different estimators. The standard deviation (SD) of the 400 parameter estimates is used as a benchmark to evaluate the average asymptotic standard errors (SE). We also reported the percentage of probability coverage ($P_{0.95}$) based on the sandwich estimate of the covariance matrix.

We used the quasi-likelihood criterion (QIC) as a model-selection criterion to choose among the different working correlation structures. The QIC is a criterion developed by Pan (2001) for model selection and selection of working correlation structures. The QIC is a modification of the Akaike Information Criterion (AIC) for the GEE model. In our case, the statistic is defined as

$$QIC(\boldsymbol{R}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{m_i}\frac{\widehat{\varepsilon}_{it}^2}{\widehat{\sigma}^2} + 2\,\text{Trace}\left(\widehat{\boldsymbol{\Omega}}_{\mathbf{I}}\widehat{\boldsymbol{V}}_{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}})\right), \tag{2.19}$$

where $\widehat{\boldsymbol{V}}_{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}})$ is the robust covariance estimate and $\widehat{\boldsymbol{\Omega}}_{\mathbf{I}}$ is the inverse of the covariance estimate under the independent working correlation evaluated at $\boldsymbol{\beta}_\tau(\boldsymbol{R})$, the parameter estimate under the working correlation of interest.

To compare with the quantile regression approach, the simulation results of the linear

quantile mixed model (LQMM) (Geraci et Bottai, 2007, 2014) were reported. The LQMM is a conditional quantile regression model with random effects parameters included to account for the within-subject dependence. The choice of the LQMM is motivated by the fact that the linear mixed model (LMM) estimate is equivalent to the exchangeable correlation estimate in the linear Gaussian setting, when $\tau = 0.5$ (Liang et Zeger, 1986). The simulation was carried out by choosing, for each distribution, the asymmetric points for which the quantiles are equal to the expectiles. For example, the Gaussian quantiles of $\tau = (0.33, 0.5, 0.67)$ correspond to the Gaussian expectiles of $\tau = (0.25, 0.5, 0.75)$.

For the sake of brevity, we present in this paper only the simulation results for the normal distribution. The simulation results for the other distributions (Student and Chi-Square) are in the **Supplementary material I**. We also published these results and the codes on our GitHub repository (https ://github.com/AmBarry/expectgee).

Table 2.1 and **Table 2.2** report the Bias and EFF results, respectively, for the (M_0) and $(M_{1/10})$ models when the error follows a multivariate normal distribution with an AR1 correlation structure and $\rho \in \{0.1, 0.5, 0.9\}$. Overall, the estimation biases are all very close to 0 for the location-shift and location-scale-shift scenarios. The bias of the three estimators is comparable for the three values $0.1, 0.5$ and 0.9 of ρ. The Un and AR1 estimators do slightly better than the Ind estimator in terms of efficiency, particularly when the correlation is higher $\rho \in (0.5, 0.9)$.

To evaluate the asymptotic standard error (SE), we use the standard deviation (SD) as a benchmark. The results are presented in **Table 2.3** and **Table 2.4** when the error follows a normal distribution respectively for the (M_0) and $(M_{1/10})$ models. We

reported in the same tables, **Table 2.3** and **Table 2.4**, the probability coverage of the different correlation structures. We observe that the values of SD and SE decrease as n becomes large. In general, the values of SD and SE are very close, particularly in the location-shift scenario. However, from time to time, the results show underestimate standard errors in the location-scale-shift scenario. This underestimation is more pronounced for the standard error of the Un correlation structure. This result is confirmed by the values of the probability coverage. The estimators of the Ind and Ar1 correlation structures have generally a very good probability coverage, even though their values sometimes fall below the nominal value (95%). On the other hand, the probability coverage of the uN estimator is, with some exceptions, below the nominal value, in the location-scale-shift scenario. Similar performances are observed when the error is generated by a Student or a Chi-Square distribution. These results can be found in the **Supplementary material I**.

Overall, the different estimators are efficient and have small biases regardless of the correlation structure. Consequently, our results confirm that the GEEE method yields a consistent and highly efficient estimator even with a misspecification of the true covariance structure (AR1).

Table 2.5 presents the QIC results of the different correlation structures with respect to the balanced/unbalanced data, the $M_0/M_{1/10}$ model and the sample sizes $n \in \{50, 100\}$. The QIC is most likely to correctly select the AR1 structure from the four given correlation structures in the M_0 scenario, particularly for the unbalanced data. In the $M_{1/10}$ scenario, the QIC is most likely to select either the AR1 or the Un structure. This last result is unexpected but not surprising. Similar results have been reported by Jang (2011).

As in Pan's paper (Pan, 2001), many studies did not include the unrestricted structure correlation as a candidate in the evaluation of the QIC or other criteria (Jang, 2011). But when included, the results showed that the QIC was strongly biased towards selecting the unrestricted structure. Please, see (Jang, 2011) and the reference therein for those studies and for further details.

The last **Tables 2.6-2.7** report the simulation results (Bias and probability coverage) of the LQMM estimator and the GEEE estimator with exchangeable working correlation. Generally, both methods are competitive in term of Bias and probability coverage (95%). Nevertheless, the LQMM model performs less well, in terms of bias and probability coverage, in the location-shift scenario, with errors generated by a Chi-square distribution.

Tableau 2.1: **Location-shift scenario** – Bias and relative efficiency of GEEE estimator with different correlation structures at 3 percentiles with $\rho \in (0.1, 0.5, 0.9)$, and the error term $\varepsilon \sim \mathcal{N}(0, 1)$.

			$m = 4$				$m \sim \mathcal{U}(3, 7)$			
			50		100		50		100	
τ	ρ	GEEE	Bias	EFF	Bias	EFF	Bias	EFF	Bias	EFF
0.25	0.1	Ind	-0.0006	1.00	-0.0003	1.00	0.0006	1.00	-0.0007	1.00
		Ar1	-0.0006	1.00	-0.0002	1.00	0.0008	0.92	-0.0006	1.00
		Un	-0.0003	0.93	-0.0001	1.00	0.0006	1.00	-0.0006	1.11
	0.5	Ind	0.0001	1.00	-0.0007	1.00	-0.0002	1.00	0.0002	1.00
		Ar1	0.0002	0.86	-0.0005	0.80	-0.0002	0.83	0.0002	0.78
		Un	-0.0003	0.79	-0.0002	0.80	-0.0003	0.92	0.0004	0.89
	0.9	Ind	-0.0009	1.00	0.0003	1.00	0.0003	1.00	0.0001	1.00
		Ar1	-0.0003	0.36	0.0000	0.40	0.0003	0.38	0.0000	0.33
		Un	-0.0007	0.71	-0.0001	0.40	0.0001	0.69	0.0005	0.67
0.50	0.1	Ind	-0.0006	1.00	0.0000	1.00	0.0004	1.00	-0.0009	1.00
		Ar1	-0.0006	1.00	0.0000	0.90	0.0005	1.00	-0.0008	0.89
		Un	-0.0006	1.00	0.0001	0.90	0.0002	1.00	-0.0008	1.00
	0.5	Ind	0.0000	1.00	-0.0005	1.00	-0.0003	1.00	0.0002	1.00
		Ar1	0.0001	0.85	-0.0004	0.89	-0.0004	0.83	0.0002	0.78
		Un	-0.0002	0.85	-0.0001	0.89	-0.0004	0.83	0.0003	0.89
	0.9	Ind	-0.0009	1.00	0.0004	1.00	0.0005	1.00	0.0001	1.00
		Ar1	-0.0002	0.38	0.0001	0.33	0.0003	0.33	0.0000	0.38
		Un	-0.0001	0.38	0.0001	0.33	0.0001	0.58	0.0002	0.38
0.75	0.1	Ind	-0.0006	1.00	0.0001	1.00	0.0001	1.00	-0.0010	1.00
		Ar1	-0.0006	1.00	0.0001	1.00	0.0002	1.00	-0.0009	1.00
		Un	-0.0007	0.93	0.0002	1.00	-0.0002	1.08	-0.0010	1.11
	0.5	Ind	0.0000	1.00	-0.0003	1.00	-0.0003	1.00	0.0001	1.00
		Ar1	-0.0001	0.86	-0.0003	0.80	-0.0005	0.83	0.0001	0.78
		Un	-0.0003	0.79	0.0000	0.80	-0.0007	0.92	0.0001	0.89
	0.9	Ind	-0.0008	1.00	0.0006	1.00	0.0006	1.00	0.0001	1.00
		Ar1	-0.0002	0.36	0.0002	0.40	0.0001	0.42	0.0000	0.33
		Un	-0.0007	0.64	0.0000	0.40	0.0005	0.67	-0.0001	0.56

Tableau 2.2: **Location-scale-shift scenario** – Bias and relative efficiency of GEEE estimator with different correlation structures at 3 percentiles with $\rho \in (0.1, 0.5, 0.9)$, and the error term $\varepsilon \sim \mathcal{N}(0,1)$.

			$m = 4$				$m \sim \mathcal{U}(3,7)$			
			50		100		50		100	
τ	ρ	GEEE	Bias	EFF	Bias	EFF	Bias	EFF	Bias	EFF
0.25	0.1	Ind	0.0013	1.00	0.0005	1.00	0.0007	1.00	0.0006	1.00
		Ar1	0.0010	1.00	0.0000	1.00	0.0003	1.00	0.0002	1.00
		Un	0.0022	0.81	0.0005	0.88	0.0014	0.75	0.0008	0.87
	0.5	Ind	-0.0022	1.00	0.0012	1.00	0.0015	1.00	0.0003	1.00
		Ar1	-0.0037	1.00	-0.0004	1.00	0.0001	1.00	-0.0013	1.00
		Un	-0.0025	0.83	-0.0006	0.88	0.0008	0.80	-0.0004	0.81
	0.9	Ind	0.0000	1.00	0.0024	1.00	-0.0010	1.00	0.0013	1.00
		Ar1	-0.0007	1.00	0.0021	1.00	-0.0015	1.00	0.0009	1.00
		Un	-0.0005	0.92	0.0026	0.89	-0.0005	0.78	0.0008	0.82
0.50	0.1	Ind	-0.0006	1.00	-0.0001	1.00	0.0001	1.00	0.0003	1.00
		Ar1	-0.0006	1.05	-0.0001	1.00	0.0001	1.00	0.0003	1.00
		Un	-0.0003	0.85	-0.0001	0.88	0.0004	0.70	0.0004	0.87
	0.5	Ind	-0.0031	1.00	0.0001	1.00	-0.0004	1.00	-0.0001	1.00
		Ar1	-0.0030	1.00	0.0001	1.00	-0.0004	1.00	-0.0001	1.07
		Un	-0.0026	0.83	-0.0002	0.88	-0.0004	0.75	0.0001	0.87
	0.9	Ind	-0.0015	1.00	0.0009	1.00	-0.0023	1.00	0.0006	1.00
		Ar1	-0.0019	1.00	0.0010	1.00	-0.0025	0.96	0.0006	1.00
		Un	-0.0026	0.83	0.0012	0.83	-0.0017	0.78	0.0003	0.82
0.75	0.1	Ind	-0.0025	1.00	-0.0009	1.00	-0.0008	1.00	0.0002	1.00
		Ar1	-0.0021	1.00	-0.0005	1.06	-0.0004	1.00	0.0006	1.00
		Un	-0.0026	0.81	-0.0009	0.88	-0.0010	0.70	0.0003	0.81
	0.5	Ind	-0.0040	1.00	-0.0009	1.00	-0.0020	1.00	-0.0006	1.00
		Ar1	-0.0024	1.00	0.0006	1.00	-0.0007	1.05	0.0009	1.00
		Un	-0.0025	0.83	0.0004	0.88	-0.0016	1.05	0.0007	0.81
	0.9	Ind	-0.0024	1.00	-0.0006	1.00	-0.0039	1.00	-0.0001	1.00
		Ar1	-0.0027	0.96	-0.0001	1.00	-0.0038	0.96	0.0004	1.00
		Un	-0.0048	0.88	-0.0002	0.89	-0.0039	0.74	-0.0002	1.06

Tableau 2.3: **Location-shift scenario** – Standard deviation, asymptotic standard errors and coverage probability of the GEEE estimator with different correlation structures at 3 percentiles with $\rho \in (0.1, 0.5, 0.9)$, and the error term $\varepsilon \sim \mathcal{N}(0,1)$.

			$m = 4$						$m \sim \mathcal{U}(3,7)$					
			50			100			50			100		
τ	ρ	GEEE	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$
0.25	0.1	Ind	0.014	0.014	0.98	0.010	0.010	0.99	0.014	0.013	0.97	0.009	0.009	0.99
		Ar1	0.014	0.014	0.97	0.010	0.010	0.99	0.013	0.012	0.98	0.009	0.009	0.99
		Un	0.014	0.013	0.97	0.010	0.010	0.98	0.014	0.013	0.97	0.009	0.010	0.99
	0.5	Ind	0.014	0.014	0.97	0.010	0.010	0.98	0.012	0.012	0.99	0.009	0.009	1.00
		Ar1	0.012	0.012	0.98	0.008	0.008	0.99	0.011	0.010	0.98	0.008	0.007	0.98
		Un	0.013	0.011	0.96	0.008	0.008	0.99	0.011	0.011	0.97	0.008	0.008	0.99
	0.9	Ind	0.014	0.014	0.97	0.009	0.010	0.99	0.014	0.013	0.97	0.009	0.009	1.00
		Ar1	0.006	0.005	0.99	0.004	0.004	1.00	0.005	0.005	0.99	0.004	0.003	1.00
		Un	0.011	0.010	0.90	0.007	0.004	0.95	0.009	0.009	0.95	0.006	0.006	0.98
0.50	0.1	Ind	0.013	0.013	0.97	0.010	0.010	0.99	0.013	0.012	0.98	0.009	0.009	0.99
		Ar1	0.013	0.013	0.98	0.010	0.009	0.98	0.013	0.012	0.97	0.009	0.008	0.99
		Un	0.014	0.013	0.97	0.010	0.009	0.99	0.013	0.012	0.97	0.009	0.009	0.98
	0.5	Ind	0.013	0.013	0.98	0.010	0.009	0.98	0.012	0.012	0.99	0.009	0.009	1.00
		Ar1	0.011	0.011	0.98	0.008	0.008	0.99	0.010	0.010	1.00	0.007	0.007	0.99
		Un	0.012	0.011	0.96	0.008	0.008	0.98	0.010	0.010	0.98	0.007	0.008	0.99
	0.9	Ind	0.013	0.013	0.98	0.009	0.009	0.99	0.013	0.012	0.98	0.009	0.008	1.00
		Ar1	0.005	0.005	0.99	0.004	0.003	1.00	0.005	0.004	0.99	0.003	0.003	1.00
		Un	0.008	0.005	0.97	0.005	0.003	0.98	0.006	0.007	0.98	0.004	0.003	0.99
0.75	0.1	Ind	0.014	0.014	0.96	0.010	0.010	0.99	0.014	0.012	0.97	0.009	0.009	0.98
		Ar1	0.014	0.014	0.96	0.010	0.010	0.99	0.014	0.012	0.96	0.009	0.009	0.98
		Un	0.015	0.013	0.95	0.010	0.010	0.98	0.014	0.013	0.96	0.009	0.010	0.99
	0.5	Ind	0.014	0.014	0.98	0.010	0.010	0.98	0.012	0.012	0.98	0.009	0.009	0.99
		Ar1	0.012	0.012	0.98	0.008	0.008	0.99	0.011	0.010	0.98	0.008	0.007	1.00
		Un	0.013	0.011	0.95	0.009	0.008	0.98	0.011	0.011	0.98	0.008	0.008	0.99
	0.9	Ind	0.014	0.014	0.97	0.010	0.010	0.98	0.014	0.012	0.98	0.009	0.009	0.99
		Ar1	0.006	0.005	0.98	0.004	0.004	1.00	0.005	0.005	0.99	0.004	0.003	1.00
		Un	0.011	0.009	0.90	0.006	0.004	0.96	0.009	0.008	0.94	0.006	0.005	0.97

Tableau 2.4: **Location-scale-shift scenario** – Standard deviation, asymptotic standard errors and coverage probability of the GEEE estimator with different correlation structures at 3 percentiles with $\rho \in (0.1, 0.5, 0.9)$, and the error term $\varepsilon \sim \mathcal{N}(0,1)$.

			$m = 4$						$m \sim \mathcal{U}(3,7)$					
			50			100			50			100		
τ	ρ	GEEE	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$
0.25	0.1	Ind	0.026	0.021	0.92	0.019	0.017	0.94	0.022	0.020	0.94	0.017	0.015	0.97
		Ar1	0.026	0.021	0.92	0.019	0.017	0.95	0.023	0.020	0.94	0.017	0.015	0.97
		Un	0.026	0.017	0.87	0.020	0.015	0.91	0.023	0.015	0.85	0.017	0.013	0.94
	0.5	Ind	0.027	0.023	0.93	0.019	0.017	0.93	0.024	0.020	0.94	0.017	0.016	0.97
		Ar1	0.028	0.023	0.92	0.020	0.017	0.94	0.025	0.020	0.94	0.018	0.016	0.97
		Un	0.029	0.019	0.84	0.020	0.015	0.90	0.025	0.016	0.86	0.018	0.013	0.93
	0.9	Ind	0.028	0.024	0.92	0.020	0.018	0.95	0.028	0.023	0.91	0.020	0.017	0.95
		Ar1	0.029	0.024	0.92	0.021	0.018	0.94	0.028	0.023	0.91	0.020	0.017	0.94
		Un	0.030	0.022	0.85	0.021	0.016	0.90	0.030	0.018	0.80	0.021	0.014	0.89
0.50	0.1	Ind	0.024	0.020	0.96	0.018	0.016	0.95	0.022	0.020	0.95	0.016	0.015	0.97
		Ar1	0.024	0.021	0.96	0.018	0.016	0.95	0.023	0.020	0.94	0.017	0.015	0.97
		Un	0.025	0.017	0.90	0.019	0.014	0.92	0.023	0.014	0.84	0.017	0.013	0.93
	0.5	Ind	0.025	0.023	0.93	0.019	0.017	0.94	0.023	0.020	0.94	0.017	0.015	0.97
		Ar1	0.026	0.023	0.92	0.020	0.017	0.94	0.024	0.020	0.94	0.017	0.016	0.97
		Un	0.027	0.019	0.87	0.020	0.015	0.90	0.024	0.015	0.84	0.017	0.013	0.92
	0.9	Ind	0.028	0.024	0.94	0.020	0.018	0.96	0.026	0.023	0.94	0.019	0.017	0.96
		Ar1	0.029	0.024	0.92	0.020	0.018	0.96	0.027	0.022	0.94	0.019	0.017	0.96
		Un	0.029	0.020	0.87	0.020	0.015	0.93	0.028	0.018	0.79	0.019	0.014	0.90
0.75	0.1	Ind	0.024	0.021	0.92	0.019	0.016	0.96	0.023	0.020	0.94	0.018	0.016	0.95
		Ar1	0.025	0.021	0.93	0.019	0.017	0.96	0.024	0.020	0.94	0.018	0.016	0.96
		Un	0.025	0.017	0.88	0.019	0.014	0.92	0.024	0.014	0.82	0.018	0.013	0.91
	0.5	Ind	0.025	0.023	0.92	0.020	0.017	0.96	0.024	0.020	0.91	0.017	0.016	0.96
		Ar1	0.026	0.023	0.92	0.021	0.017	0.94	0.024	0.021	0.92	0.018	0.016	0.96
		Un	0.027	0.019	0.86	0.021	0.015	0.92	0.025	0.021	0.83	0.018	0.013	0.92
	0.9	Ind	0.030	0.025	0.92	0.020	0.018	0.96	0.027	0.023	0.94	0.019	0.017	0.96
		Ar1	0.030	0.024	0.91	0.020	0.018	0.95	0.027	0.022	0.92	0.019	0.017	0.96
		Un	0.032	0.022	0.84	0.021	0.016	0.92	0.029	0.017	0.82	0.020	0.018	0.90

Tableau 2.5: Total percentage of the working correlation matrix selected by QIC for the different correlations $\rho \in \{0.1, 0.5, 0.9\}$ from 1200 independent replications. The true correlation matrix is AR1.

	Balanced Data								Unbalanced Data							
	$n = 50$				$n = 100$				$n = 50$				$n = 100$			
	Ind	Exc	AR1	Un	Ind	Exc	AR1	Un	Ind	Exc	AR1	Un	Ind	Exc	AR1	Un
	Location-shift model															
$\mathcal{N}$	3.58	6.83	45.83	43.75	3.42	5.08	48.42	43.08	3.83	8.25	63.25	24.67	2.67	4.08	55.42	37.83
$\mathcal{T}_3$	8.17	14.00	46.67	31.17	8.42	11.83	48.92	30.83	8.00	14.08	59.17	18.75	5.83	11.25	58.50	24.42
χ^3_2	5.75	11.50	45.67	37.08	5.00	9.42	48.08	37.50	4.00	10.83	61.92	23.25	4.08	6.58	58.67	30.67
	Location-scale-shift model															
$\mathcal{N}$	16.75	16.67	20.42	46.17	19.33	15.42	23.25	42.00	13.67	18.25	27.08	41.00	13.17	16.17	23.67	47.00
$\mathcal{T}_3$	8.25	13.67	43.25	34.83	10.75	14.75	38.83	35.67	10.42	11.92	39.50	38.17	10.33	14.33	36.83	38.50
χ^3_2	4.25	9.58	37.92	48.25	3.42	6.83	38.17	51.67	3.50	9.33	39.00	48.17	4.42	7.00	40.42	48.17

Tableau 2.6: Bias and probability coverage of the GEEE estimator with exchangeable working correlation and the LQMM estimator at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ for a **balanced** panel with the error term $\varepsilon \sim \mathcal{N}(0,1)$.

		$n = 50$				$n = 100$			
		Bias		$P_{0.95}$		Bias		$P_{0.95}$	
ρ	τ	Exc	LQMM	Exc	LQMM	Exc	LQMM	Exc	LQMM
				Location-shift model					
0.1	τ_1	-0.0005	-0.0008	0.97	1.00	-0.0002	-0.0001	0.98	0.99
	τ_2	-0.0006	-0.0008	0.98	0.99	0.0000	0.0000	0.99	0.98
	τ_3	-0.0006	-0.0008	0.97	0.98	0.0001	0.0001	0.99	0.99
0.5	τ_1	0.0005	0.0000	0.97	0.99	-0.0007	0.0000	0.98	0.99
	τ_2	0.0003	0.0007	0.97	0.98	-0.0005	-0.0005	0.99	0.99
	τ_3	0.0001	0.0000	0.98	0.98	-0.0004	-0.0002	0.99	0.99
0.9	τ_1	0.0001	0.0002	0.99	0.99	0.0003	-0.0003	0.99	0.99
	τ_2	0.0001	0.0000	1.00	0.99	0.0004	0.0005	1.00	0.99
	τ_3	0.0002	-0.0001	0.98	0.98	0.0005	0.0008	0.99	1.00
				Location-scale-shift model					
0.1	τ_1	-0.0008	-0.0003	0.99	0.99	-0.0007	0.0007	0.99	0.98
	τ_2	-0.0001	-0.0005	0.99	1.00	0.0000	0.0001	1.00	0.99
	τ_3	0.0006	0.0000	0.98	0.99	0.0008	0.0005	0.98	0.98
0.5	τ_1	-0.0019	0.0008	0.98	0.99	-0.0018	0.0002	0.97	0.99
	τ_2	-0.0002	0.0008	1.00	0.98	-0.0001	-0.0002	1.00	0.98
	τ_3	0.0015	0.0006	0.98	0.99	0.0016	0.0002	0.98	0.98
0.9	τ_1	-0.0018	0.0004	0.97	0.99	-0.0017	0.0000	0.99	1.00
	τ_2	-0.0001	-0.0003	0.98	0.99	0.0000	-0.0004	1.00	0.99
	τ_3	0.0017	-0.0005	0.98	0.99	0.0017	-0.0008	0.97	0.99

Tableau 2.7: Bias and probability coverage of the GEEE estimator with exchangeable working correlation and the LQMM estimator at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ for an **unbalanced** panel with the error term $\varepsilon \sim \mathcal{N}(0,1)$.

		$n = 50$				$n = 100$			
		Bias		$P_{0.95}$		Bias		$P_{0.95}$	
ρ	τ	Exc	LQMM	Exc	LQMM	Exc	LQMM	Exc	LQMM
				Location-shift model					
0.1	τ_1	0.0007	0.0010	0.97	0.97	-0.0006	-0.0006	0.99	0.99
	τ_2	0.0005	0.0009	0.97	0.99	-0.0008	-0.0009	0.99	1.00
	τ_3	0.0001	0.0001	0.97	0.97	-0.0009	-0.0009	0.98	1.00
0.5	τ_1	-0.0002	-0.0003	0.98	0.99	0.0002	0.0001	1.00	0.99
	τ_2	-0.0003	-0.0006	0.99	0.99	0.0002	0.0001	1.00	0.99
	τ_3	-0.0004	0.0004	0.97	0.98	0.0002	0.0000	1.00	1.00
0.9	τ_1	0.0005	0.0009	0.99	0.98	0.0000	0.0003	1.00	1.00
	τ_2	0.0003	0.0007	0.99	0.99	0.0000	-0.0006	1.00	1.00
	τ_3	0.0000	0.0003	0.99	1.00	0.0000	0.0003	1.00	1.00
				Location-scale-shift model					
0.1	τ_1	-0.0007	-0.0015	0.99	0.99	-0.0008	0.0007	1.00	0.99
	τ_2	0.0000	-0.0014	0.99	0.99	-0.0001	0.0011	1.00	0.99
	τ_3	0.0006	-0.0007	0.98	0.98	0.0005	0.0009	1.00	1.00
0.5	τ_1	-0.0015	-0.0006	0.99	0.98	-0.0015	0.0013	0.99	0.99
	τ_2	-0.0001	0.0001	1.00	0.98	0.0000	0.0008	1.00	0.98
	τ_3	0.0013	-0.0009	0.99	0.99	0.0014	-0.0003	0.99	0.99
0.9	τ_1	-0.0017	0.0004	0.97	0.98	-0.0017	-0.0002	0.99	1.00
	τ_2	0.0000	-0.0005	0.98	0.98	0.0000	0.0004	1.00	1.00
	τ_3	0.0017	0.0008	0.97	0.97	0.0016	0.0002	0.98	0.99

2.5 Application

In this Section, the proposed method is applied to the repeated measurements labor pain dataset previously reported by Davis (1991). It is a commonly used dataset in biostatistics, and used several times in the quantile regression framework (Jung, 1996; Geraci et Bottai, 2007; Lu et Fan, 2015). The dataset comes from a clinical trial on the effectiveness of a medication for labor pain relief for 83 women in labor. The treatment group (43 women) and the placebo group (40 women) were randomly assigned. The response variable is a self-reported score measured every 30 min on a 100-mm line, where 0 means no pain and 100 means extreme pain. A nearly monotone pattern of missing data was found for the response variable, and the maximum number of measurements per woman was six. **Figure 2.1** shows the box-plot of the response variable for all women by treatment group. At first glance, we can determine that the response variable is non-normal. We also observe the evolution of the mean and the median over time. The objective is to study the effect of medication on the self-reported pain score from the following equation

$$y_{it} = \beta_0 + \beta_1 R_i + \beta_2 T_{it} + \beta_3 R_i T_{it} + \varepsilon_{it}, \tag{2.20}$$

where y_{it} is the t-th measure of the pain for the i-th subject. R_i is the treatment variable with value 0 for placebo and 1 for treatment, and T_{it} is the measurement time divided by 30min. The corresponding GEEE equation, for a fixed τ

$$\mu_\tau(y_{it}) = \beta_{0\tau} + \beta_1 R_i + \beta_2 T_{it} + \beta_3 R_i T_{it}, \tag{2.21}$$

with four working correlations (Ind, Exc, AR1, Un), was estimated for three asymmetric points, $(0.25, 0.5, 0.75)$. Note that, in the presence of heteroscedasticity, the random error is correlated to the covariates and as in the simulation Section both

parameters, the intercept and the slopes, will vary with τ. **Table 2.8** presents the results of the estimated parameters, their standard errors, as well as their 95% confidence intervals. It is observed that the different GEEE methods produce comparable estimates. The estimated parameter $\widehat{\beta}_0$ is not significant except for the percentile $\tau = 0.75$. This indicates that the baseline pain does not differ significantly between the two groups. The estimated parameters $\widehat{\beta}_2$ and $\widehat{\beta}_3$ are significant at 5% level for the GEEE methods, except for $\widehat{\beta}_3$ of the GEEE with an Exc working correlation. This means that time and its interaction with treatment affects the level of pain. To investigate the effect of the drug on pain over time, we study the evolution of this difference

$$\mu_\tau(y_{it}|R_i = 1) - \mu_\tau(y_{it}|R_i = 0) = \beta_1 + \beta_3 T_{it},$$

for which the result is presented in **Figure 2.2**. We see that medication helps women feel a relief from their labor pain. Indeed, the pain of women in the placebo group grows faster with time than that of the treated group, and this is for all the GEEE methods and at different percentiles $(0.25, 0.5, 0.75)$. We also observe that the difference between the placebo group and the treated group varies according to the different values of τ.

Using the QIC measure to choose among the 4 working correlation structures leads us to select the Un ($QIC = 2416.515$) or the AR1 ($QIC = 2416.924$) correlation structures over the Exc ($QIC = 2418.182$) or the Ind ($QIC = 2419.414$) structures. These results are consistent with the structure of the dataset. The repeated data are uniformly spaced in time and the correlation of the response variable is stronger for adjacent measurements than for distant ones.

This real application shows that the GEEE method can be an excellent complement to the GEE method which remains a widely used method for the analysis of

longitudinal data. In addition to taking into account the heterogeneity of covariate effects and the unobserved heterogeneity, the GEEE method inherits all the favorable properties of the GEE method.

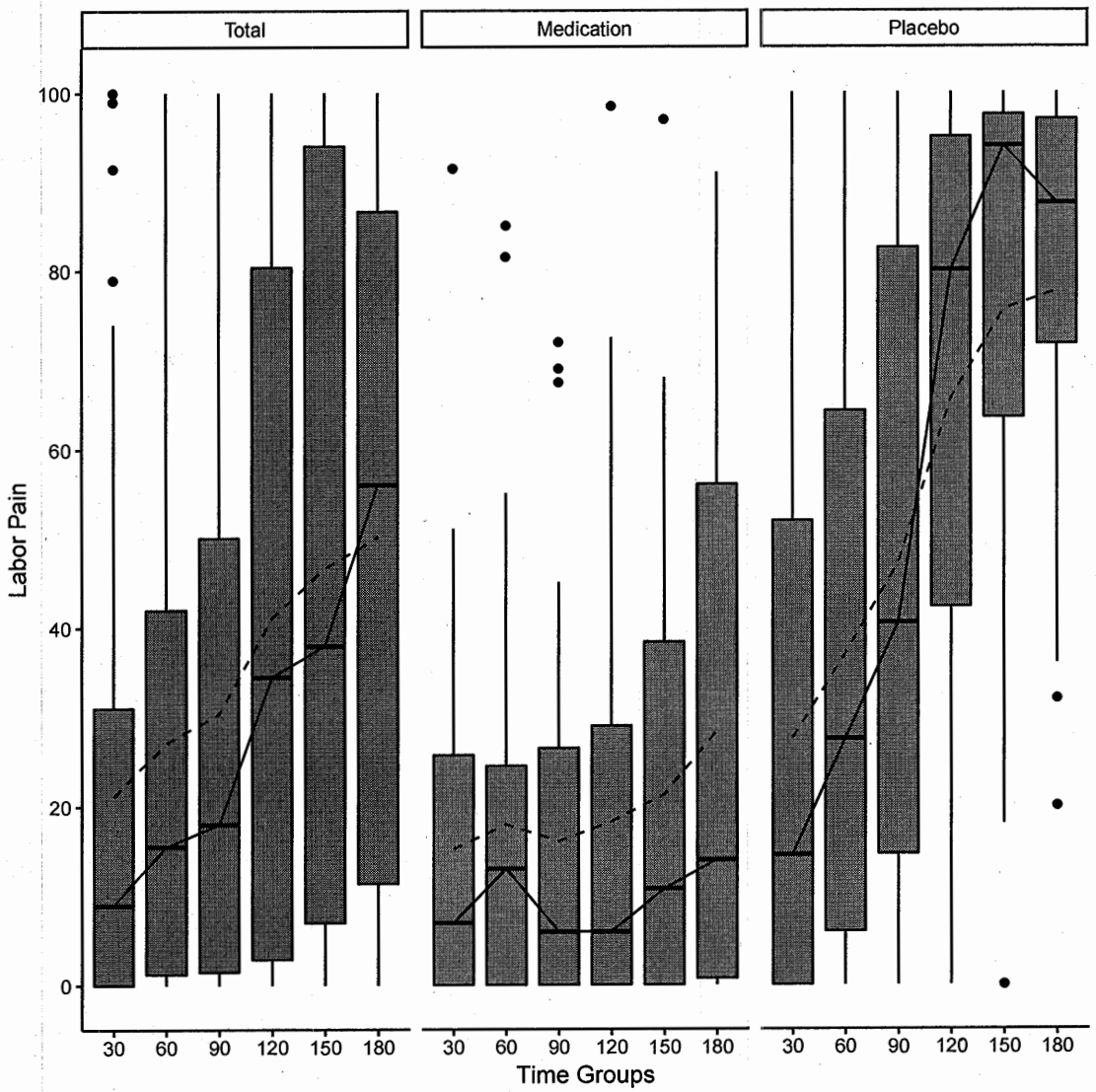

Figure 2.1: Box-plot of measured labor pain for all women in placebo and medication groups. The solid lines connect the medians and the dashed lines connect the means.

Tableau 2.8: Parameters estimates (Est) with their standard errors (SE) and 95% confidence intervals (CI) obtained from the GEEE independent, exchangeable, AR1 and unstructured working correlation at three percentiles, $\tau = 0.25, 0.5, 0.75$.

	$\tau =$	0.25				0.5				0.75			
GEEE		Est	SE	P	CI	Est	SE	P	CI	Est	SE	P	CI
Ind	β_0	2.63	4.83	0.59	(-6.83,12.09)	15.66	6.62	0.02	(2.68,28.64)	35.76	8.06	0.00	(19.97,51.55)
	β_1	4.34	5.37	0.42	(-6.19,14.86)	-2.23	7.69	0.77	(-17.31,12.85)	-12.92	9.89	0.19	(-32.30, 6.46)
	β_2	10.70	1.97	0.00	(6.83,14.57)	11.33	1.62	0.00	(8.16,14.49)	9.84	1.48	0.00	(6.94,12.74)
	β_3	-9.65	2.12	0.00	(-13.80,-5.49)	-9.58	2.03	0.00	(-13.56,-5.59)	-7.32	2.22	0.00	(-11.67,-2.97)
Exc	β_0	1.69	11.47	0.88	(-20.78,24.17)	17.14	12.61	0.17	(-7.59,41.86)	38.17	12.06	0.00	(14.53,61.81)
	β_1	1.68	12.50	0.89	(-22.83,26.18)	-4.31	14.56	0.77	(-32.85,24.22)	-12.19	14.22	0.39	(-40.07,15.69)
	β_2	10.15	4.65	0.03	(1.03,19.26)	11.25	4.31	0.01	(2.80,19.70)	10.05	3.61	0.01	(2.97,17.14)
	β_3	-8.71	4.94	0.08	(-18.39, 0.96)	-9.15	4.96	0.06	(-18.87, 0.56)	-7.03	4.34	0.11	(-15.53, 1.47)
AR1	β_0	3.53	6.37	0.58	(-8.95,16.02)	15.80	8.92	0.08	(-1.68,33.28)	33.07	10.98	0.00	(11.55,54.59)
	β_1	4.29	7.44	0.56	(-10.28,18.87)	-1.69	10.83	0.88	(-22.92,19.54)	-13.22	14.85	0.37	(-42.33,15.89)
	β_2	8.06	1.37	0.00	(5.37,10.74)	9.08	1.57	0.00	(6.01,12.16)	8.76	1.59	0.00	(5.65,11.87)
	β_3	-7.62	1.64	0.00	(-10.84,-4.40)	-8.14	2.21	0.00	(-12.46,-3.81)	-6.52	2.91	0.03	(-12.23,-0.82)
Un	β_0	2.24	6.36	0.73	(-10.23,14.70)	16.95	7.95	0.03	(1.36,32.54)	38.67	9.01	0.00	(21.00,56.34)
	β_1	2.50	6.96	0.72	(-11.15,16.14)	-4.15	9.14	0.65	(-22.07,13.76)	-12.36	10.56	0.24	(-33.06, 8.33)
	β_2	10.39	2.81	0.00	(4.88,15.89)	11.30	2.52	0.00	(6.36,16.25)	10.05	2.24	0.00	(5.65,14.45)
	β_3	-9.13	2.97	0.00	(-14.94,-3.32)	-9.21	2.97	0.00	(-15.03,-3.39)	-6.95	2.94	0.02	(-12.70,-1.19)

2.6 Conclusion

We combined weighted asymmetric least squares regression and generalized estimating equations to derive a new class of estimators : generalized expectile estimating equations estimators. This new GEEE class models the underlying correlation structure from one subject by formally including a hypothesized structure with the within-subject correlation. In addition, this new model captures the heterogeneity of covariate effects and accounts for unobserved heterogeneity. We also showed how to extend and adapt some of the most common and popular GEE working correlations.
We derived the asymptotic properties of this new estimator and proposed a robust estimator of its variance-covariance matrix. The results of the exhaustive simulations displayed its favorable qualities under various scenarios and its advantages in relation to existing methods. The QIC is most likely to select the correct working correlation (AR1) among the four working correlation structures used in the simulation. Finally, we fit the GEEE model to the labor pain data. The results revealed a strong association of treatment and time on the labor pain of the two groups of women. This result is consistent with the results obtained in other studies (Lu et Fan, 2015; Leng et Zhang, 2014). In addition, the results show that the heterogeneity of the evolution of pain according to time depends on whether one is in the center or on the left/right of the tail of the distribution response. The application of the QIC criterion to choose between the four correlation structures leads to the selection of either the AR1 or the Un working correlation structures.
The proposed model opens the door to other avenues of research. Unlike the quantile regression model, the expectile regression and the GEEE method will naturally generalize to the dichotomous or count data, or to other longitudinal models already used to estimate the effect of covariates on the average of the response.

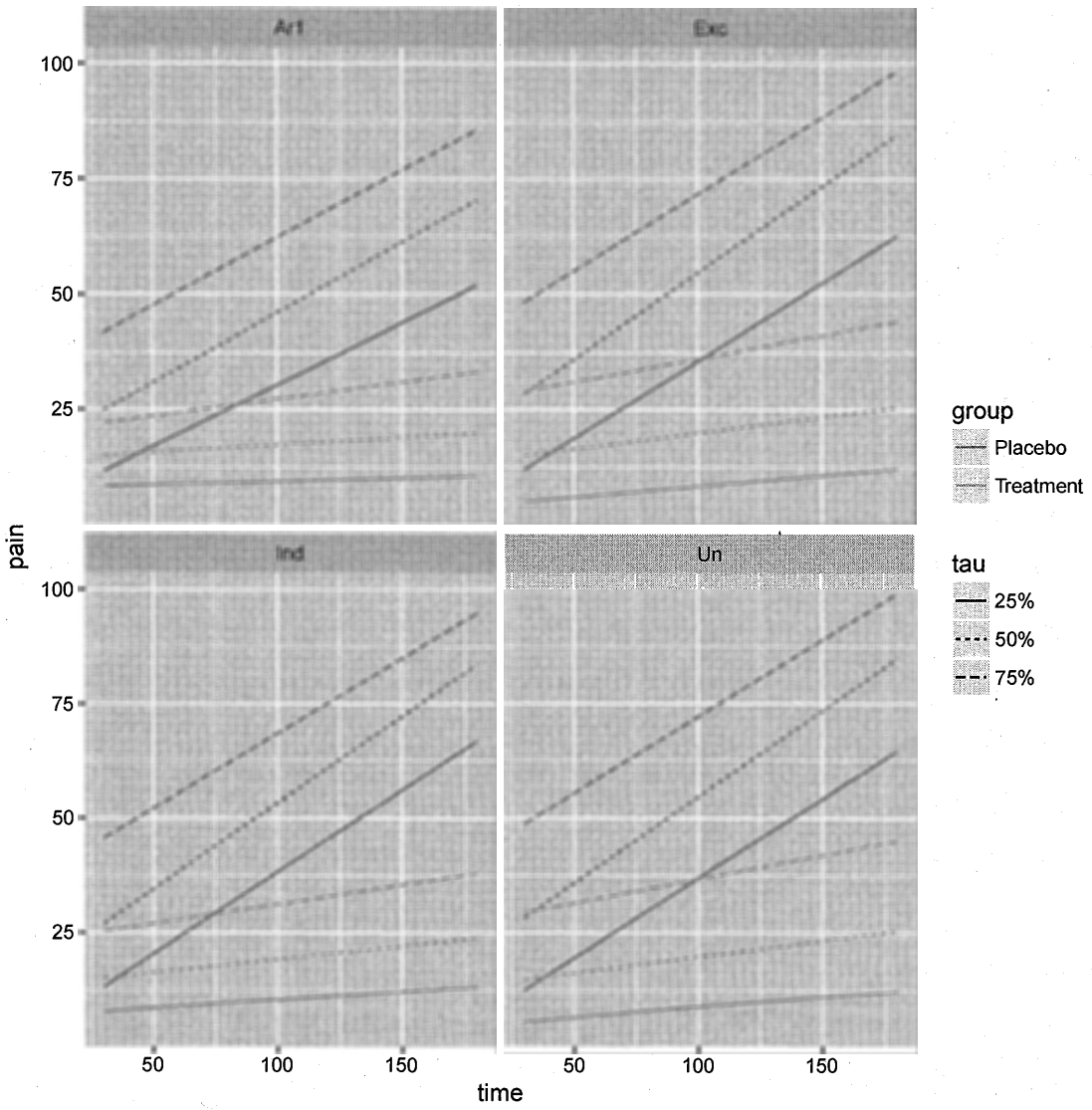

Figure 2.2: Representation of the estimated labor pain obtained from the GEEE method with the different working correlations (Ind, Exc, AR1, Un) at three percentiles, $(0.25, 0.5, 0.75)$.

2.7 Appendix

2.7.1 Supplementary Material I : Additional results

Tableau 2.9: **Location-shift scenario** – Bias and relative efficiency of GEEE estimator with different correlation structures at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ and the error term $\varepsilon \sim t_3$.

			$T = 4$				$T \sim \mathcal{U}(3,7)$			
			50		100		50		100	
τ	ρ	GEEE	Bias	EFF	Bias	EFF	Bias	EFF	Bias	EFF
0.25	0.1	Ind	0.0001	1.00	0.0023	1.00	-0.0014	1.00	0.0005	1.00
		Ar1	0.0001	1.00	0.0021	1.00	-0.0015	1.00	0.0005	1.00
		Un	0.0006	0.92	0.0024	0.89	-0.0018	1.00	0.0005	1.00
	0.5	Ind	-0.0012	1.00	0.0000	1.00	0.0001	1.00	0.0016	1.00
		Ar1	-0.0010	0.88	-0.0007	0.89	0.0011	0.88	0.0014	0.88
		Un	-0.0005	0.85	-0.0012	0.89	0.0012	1.08	0.0014	0.88
	0.9	Ind	-0.0020	1.00	-0.0002	1.00	-0.0016	1.00	-0.0004	1.00
		Ar1	-0.0002	0.44	-0.0002	0.44	-0.0009	0.48	-0.0001	0.47
		Un	0.0019	1.00	-0.0007	0.83	-0.0026	1.86	0.0004	1.76
0.50	0.1	Ind	0.0002	1.00	0.0017	1.00	-0.0011	1.00	0.0003	1.00
		Ar1	0.0003	0.95	0.0015	1.00	-0.0012	1.00	0.0004	1.00
		Un	0.0006	0.91	0.0019	0.94	-0.0012	1.05	0.0005	1.07
	0.5	Ind	-0.0002	1.00	0.0003	1.00	0.0003	1.00	0.0009	1.00
		Ar1	0.0001	0.86	-0.0006	0.81	0.0010	0.86	0.0006	0.86
		Un	0.0004	0.86	-0.0006	0.81	0.0004	0.81	0.0004	0.93
	0.9	Ind	-0.0016	1.00	-0.0004	1.00	-0.0007	1.00	0.0001	1.00
		Ar1	-0.0004	0.41	-0.0003	0.38	-0.0003	0.42	0.0002	0.43
		Un	0.0015	0.68	-0.0012	0.56	-0.0002	1.00	0.0008	0.57
0.75	0.1	Ind	0.0004	1.00	0.0017	1.00	-0.0009	1.00	-0.0002	1.00
		Ar1	0.0005	1.00	0.0016	1.00	-0.0009	1.00	-0.0002	1.00
		Un	0.0009	0.88	0.0019	0.89	-0.0007	0.96	0.0000	1.00
	0.5	Ind	0.0004	1.00	0.0004	1.00	0.0007	1.00	0.0004	1.00
		Ar1	0.0005	0.88	-0.0003	0.89	0.0013	0.88	0.0000	0.88
		Un	0.0005	0.81	0.0001	0.79	-0.0006	0.88	-0.0003	0.94
	0.9	Ind	-0.0016	1.00	-0.0006	1.00	0.0002	1.00	0.0006	1.00
		Ar1	-0.0005	0.46	-0.0004	0.42	0.0003	0.45	0.0008	0.44
		Un	0.0000	1.88	-0.0010	0.68	-0.0001	0.77	0.0001	1.00

Tableau 2.10: **Location-shift scenario** – Bias and relative efficiency of GEEE estimator with different correlation structures at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ and the error term $\varepsilon \sim \chi^2_3$.

			$T = 4$				$T \sim \mathcal{U}(3,7)$			
			50		100		50		100	
τ	ρ	GEEE	Bias	EFF	Bias	EFF	Bias	EFF	Bias	EFF
0.25	0.1	Ind	-0.0002	1.00	-0.0003	1.00	-0.0006	1.00	-0.0013	1.00
		Ar1	-0.0006	1.00	-0.0002	1.00	-0.0007	0.95	-0.0014	1.00
		Un	-0.0006	0.96	-0.0002	1.00	-0.0009	1.00	-0.0012	1.13
	0.5	Ind	-0.0007	1.00	0.0002	1.00	-0.0002	1.00	0.0010	1.00
		Ar1	-0.0012	0.83	0.0000	0.82	0.0008	0.81	0.0011	0.80
		Un	-0.0014	0.79	0.0001	0.82	0.0002	0.86	0.0013	0.93
	0.9	Ind	-0.0008	1.00	-0.0001	1.00	-0.0017	1.00	0.0006	1.00
		Ar1	-0.0005	0.37	0.0003	0.35	-0.0005	0.36	0.0004	0.40
		Un	0.0005	0.62	0.0004	0.41	-0.0011	0.50	0.0011	0.47
0.50	0.1	Ind	-0.0013	1.00	-0.0005	1.00	-0.0008	1.00	-0.0015	1.00
		Ar1	-0.0017	1.00	-0.0005	1.00	-0.0009	1.00	-0.0016	1.00
		Un	-0.0014	0.94	-0.0008	1.00	-0.0008	1.00	-0.0015	1.10
	0.5	Ind	-0.0011	1.00	0.0006	1.00	0.0001	1.00	0.0009	1.00
		Ar1	-0.0022	0.82	-0.0001	0.87	0.0014	0.83	0.0013	0.81
		Un	-0.0025	0.79	-0.0002	0.83	-0.0004	0.83	0.0016	0.90
	0.9	Ind	-0.0006	1.00	0.0001	1.00	-0.0018	1.00	0.0006	1.00
		Ar1	-0.0008	0.41	0.0003	0.39	-0.0003	0.38	0.0007	0.40
		Un	0.0016	0.56	0.0003	0.43	-0.0016	1.45	0.0008	1.05
0.75	0.1	Ind	-0.0023	1.00	-0.0007	1.00	-0.0007	1.00	-0.0018	1.00
		Ar1	-0.0028	1.00	-0.0008	1.00	-0.0008	0.98	-0.0018	1.00
		Un	-0.0024	0.91	-0.0012	0.94	0.0002	0.98	-0.0018	1.07
	0.5	Ind	-0.0015	1.00	0.0015	1.00	0.0005	1.00	0.0007	1.00
		Ar1	-0.0034	0.87	0.0001	0.88	0.0017	0.88	0.0015	0.86
		Un	-0.0043	0.82	-0.0004	0.82	-0.0021	0.90	0.0023	0.93
	0.9	Ind	-0.0007	1.00	0.0007	1.00	-0.0018	1.00	0.0007	1.00
		Ar1	-0.0012	0.43	0.0004	0.44	-0.0001	0.44	0.0008	0.41
		Un	0.0022	1.36	-0.0001	0.56	-0.0013	0.90	0.0020	1.03

Tableau 2.11: **Location-scale-shift scenario** – Bias and relative efficiency of GEEE estimator with different correlation structures at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ and the error term $\varepsilon \sim t_3$.

			$T = 4$				$T \sim \mathcal{U}(3,7)$			
			50		100		50		100	
τ	ρ	GEEE	Bias	EFF	Bias	EFF	Bias	EFF	Bias	EFF
0.25	0.1	Ind	0.0020	1.00	0.0004	1.00	0.0048	1.00	-0.0002	1.00
		Ar1	0.0014	1.03	-0.0001	1.00	0.0043	1.00	-0.0008	1.00
		Un	0.0035	0.75	0.0006	0.79	0.0048	0.68	-0.0002	0.77
	0.5	Ind	0.0033	1.00	0.0037	1.00	-0.0002	1.00	0.0008	1.00
		Ar1	0.0015	1.00	0.0018	1.04	-0.0016	1.00	-0.0009	1.00
		Un	0.0042	0.74	0.0021	0.82	-0.0011	0.63	-0.0008	0.75
	0.9	Ind	0.0000	1.00	0.0016	1.00	0.0002	1.00	-0.0014	1.00
		Ar1	-0.0003	0.98	0.0010	1.00	-0.0006	1.00	-0.0015	0.97
		Un	-0.0009	0.73	0.0017	0.80	0.0000	0.74	-0.0019	0.78
0.50	0.1	Ind	0.0010	1.00	-0.0001	1.00	0.0013	1.00	-0.0008	1.00
		Ar1	0.0010	1.00	-0.0001	1.00	0.0013	1.00	-0.0008	1.00
		Un	0.0021	0.76	0.0002	0.84	0.0017	0.71	-0.0008	0.74
	0.5	Ind	0.0025	1.00	0.0020	1.00	0.0010	1.00	0.0006	1.00
		Ar1	0.0023	1.03	0.0018	1.00	0.0011	1.00	0.0006	1.04
		Un	0.0034	0.76	0.0017	0.81	0.0006	0.62	0.0002	0.79
	0.9	Ind	-0.0012	1.00	-0.0001	1.00	-0.0022	1.00	-0.0011	1.00
		Ar1	-0.0019	0.97	-0.0007	1.00	-0.0027	0.97	-0.0010	0.96
		Un	-0.0029	0.75	-0.0006	0.81	-0.0032	0.82	-0.0022	0.74
0.75	0.1	Ind	0.0002	1.00	0.0002	1.00	-0.0013	1.00	-0.0016	1.00
		Ar1	0.0007	1.00	0.0009	1.00	-0.0008	1.00	-0.0012	1.00
		Un	0.0009	0.73	0.0007	0.77	-0.0009	0.84	-0.0017	0.77
	0,5	Ind	0.0017	1.00	0.0012	1.00	0.0025	1.00	0.0008	1.00
		Ar1	0.0031	1.00	0.0027	1.03	0.0042	1.00	0.0024	1.00
		Un	0.0026	0.76	0.0025	0.77	0.0029	0.62	0.0008	3.38
	0.9	Ind	-0.0031	1.00	-0.0011	1.00	-0.0046	1.00	-0.0019	1.00
		Ar1	-0.0041	0.97	-0.0017	0.97	-0.0050	0.97	-0.0019	0.97
		Un	-0.0079	0.87	-0.0013	0.78	-0.0069	1.51	-0.0044	0.73

Tableau 2.12: **Location-scale-shift scenario** – Bias and relative efficiency of GEEE estimator with different correlation structures at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ and the error $\varepsilon \sim \chi^2_3$.

			$T = 4$				$T \sim \mathcal{U}(3,7)$			
			50		100		50		100	
τ	ρ	GEEE	Bias	EFF	Bias	EFF	Bias	EFF	Bias	EFF
0.25	0.1	Ind	0.0012	1.00	0.0049	1.00	-0.0023	1.00	0.0039	1.00
		Ar1	0.0005	1.00	0.0038	1.00	-0.0033	1.00	0.0030	1.00
		Un	0.0032	0.84	0.0047	0.90	-0.0007	0.74	0.0042	0.88
	0.5	Ind	0.0018	1.00	0.0010	1.00	0.0026	1.00	0.0023	1.00
		Ar1	-0.0019	1.00	-0.0025	1.03	-0.0005	1.03	-0.0010	1.04
		Un	-0.0008	0.85	-0.0028	0.90	0.0015	0.75	-0.0008	0.89
	0.9	Ind	-0.0009	1.00	-0.0001	1.00	0.0034	1.00	0.0018	1.00
		Ar1	-0.0013	1.00	-0.0011	1.00	0.0019	1.00	0.0005	1.00
		Un	-0.0011	0.83	-0.0015	0.88	0.0030	0.78	0.0001	0.87
0.50	0.1	Ind	-0.0023	1.00	0.0052	1.00	-0.0049	1.00	0.0035	1.00
		Ar1	-0.0022	1.00	0.0051	1.00	-0.0051	1.00	0.0036	1.03
		Un	-0.0006	0.82	0.0051	0.85	-0.0039	0.73	0.0039	0.83
	0.5	Ind	-0.0009	1.00	-0.0003	1.00	-0.0003	1.00	0.0013	1.00
		Ar1	-0.0018	1.02	-0.0007	1.02	-0.0006	1.02	0.0010	1.03
		Un	-0.0022	0.82	-0.0013	0.88	0.0002	0.71	-0.0002	0.81
	0.9	Ind	-0.0040	1.00	-0.0018	1.00	0.0022	1.00	-0.0002	1.00
		Ar1	-0.0035	1.00	-0.0018	1.00	0.0012	0.98	-0.0008	1.00
		Un	-0.0049	0.84	-0.0028	0.86	0.0011	0.85	-0.0012	0.90
0.75	0.1	Ind	-0.0069	1.00	0.0041	1.00	-0.0097	1.00	0.0031	1.00
		Ar1	-0.0059	1.00	0.0051	1.02	-0.0090	1.00	0.0040	1.00
		Un	-0.0058	0.78	0.0041	0.81	-0.0098	0.79	0.0034	0.81
	0.5	Ind	-0.0055	1.00	-0.0031	1.00	-0.0048	1.00	-0.0011	1.00
		Ar1	-0.0028	1.01	0.0005	1.02	-0.0017	1.02	0.0025	1.02
		Un	-0.0046	0.88	-0.0008	0.81	-0.0021	0.71	0.0015	0.82
	0.9	Ind	-0.0092	1.00	-0.0059	1.00	-0.0013	1.00	-0.0032	1.00
		Ar1	-0.0077	0.97	-0.0044	1.00	-0.0017	0.97	-0.0028	0.98
		Un	-0.0111	0.84	-0.0059	1.14	-0.0026	1.32	-0.0036	1.02

Tableau 2.13: **Location-shift scenario** – Standard deviation, asymptotic standard errors and probability coverage of the GEEE estimator with different correlation structures at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ and the error term $\varepsilon \sim t_3$.

			$m = 4$						$m \sim \mathcal{U}(3,7)$					
			50			100			50			100		
τ	ρ	GEEE	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$
0.25	0.1	Ind	0.026	0.025	0.96	0.019	0.019	0.98	0.024	0.023	0.97	0.018	0.017	0.98
		Ar1	0.026	0.025	0.97	0.019	0.019	0.98	0.024	0.023	0.97	0.018	0.017	0.98
		Un	0.027	0.023	0.93	0.020	0.017	0.95	0.029	0.023	0.94	0.018	0.017	0.97
	0.5	Ind	0.029	0.026	0.96	0.019	0.018	0.98	0.026	0.024	0.97	0.018	0.017	0.98
		Ar1	0.026	0.023	0.98	0.017	0.016	0.98	0.024	0.021	0.97	0.017	0.015	0.98
		Un	0.033	0.022	0.89	0.020	0.016	0.94	0.029	0.026	0.89	0.019	0.015	0.93
	0.9	Ind	0.029	0.025	0.96	0.020	0.018	0.98	0.026	0.021	0.96	0.017	0.017	0.99
		Ar1	0.014	0.011	0.98	0.009	0.008	0.98	0.012	0.010	0.99	0.010	0.008	0.99
		Un	0.031	0.025	0.80	0.019	0.015	0.83	0.020	0.039	0.87	0.016	0.030	0.87
0.50	0.1	Ind	0.022	0.022	0.97	0.016	0.016	0.98	0.020	0.019	0.98	0.015	0.014	0.98
		Ar1	0.022	0.021	0.97	0.016	0.016	0.98	0.019	0.019	0.98	0.015	0.014	0.99
		Un	0.023	0.020	0.95	0.016	0.015	0.96	0.021	0.020	0.96	0.015	0.015	0.98
	0.5	Ind	0.024	0.022	0.98	0.016	0.016	0.98	0.023	0.021	0.98	0.015	0.014	0.98
		Ar1	0.022	0.019	0.97	0.014	0.013	0.99	0.021	0.018	0.97	0.013	0.012	0.98
		Un	0.025	0.019	0.93	0.015	0.013	0.96	0.024	0.017	0.93	0.014	0.013	0.97
	0.9	Ind	0.023	0.022	0.98	0.017	0.016	0.99	0.021	0.019	0.96	0.014	0.014	1.00
		Ar1	0.010	0.009	0.99	0.007	0.006	0.99	0.009	0.008	0.99	0.007	0.006	1.00
		Un	0.020	0.015	0.85	0.012	0.009	0.90	0.015	0.019	0.91	0.011	0.008	0.94
0.75	0.1	Ind	0.027	0.025	0.97	0.020	0.019	0.97	0.023	0.023	0.98	0.018	0.017	0.98
		Ar1	0.027	0.025	0.97	0.020	0.019	0.97	0.022	0.023	0.98	0.018	0.017	0.98
		Un	0.029	0.022	0.94	0.020	0.017	0.96	0.026	0.022	0.94	0.018	0.017	0.96
	0.5	Ind	0.030	0.026	0.96	0.020	0.019	0.98	0.030	0.024	0.98	0.018	0.017	0.98
		Ar1	0.028	0.023	0.97	0.018	0.017	0.98	0.028	0.021	0.98	0.016	0.015	0.98
		Un	0.033	0.021	0.89	0.020	0.015	0.93	0.035	0.021	0.90	0.017	0.016	0.94
	0.9	Ind	0.026	0.024	0.98	0.021	0.019	0.98	0.025	0.022	0.96	0.016	0.016	1.00
		Ar1	0.013	0.011	0.98	0.010	0.008	0.99	0.011	0.010	0.98	0.008	0.007	1.00
		Un	0.031	0.045	0.78	0.018	0.013	0.84	0.020	0.017	0.89	0.015	0.016	0.88

Tableau 2.14: **Location-shift scenario** – Standard deviation, asymptotic standard errors and probability coverage of the GEEE estimator with different correlation structures at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ and the error term $\varepsilon \sim \chi^2_3$.

			$m = 4$						$m \sim \mathcal{U}(3,7)$					
			50			100			50			100		
τ	ρ	GEEE	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$
0.25	0.1	Ind	0.025	0.024	0.97	0.019	0.017	0.96	0.022	0.022	0.96	0.016	0.015	0.97
		Ar1	0.025	0.024	0.96	0.019	0.017	0.96	0.022	0.021	0.96	0.016	0.015	0.98
		Un	0.026	0.023	0.94	0.020	0.017	0.95	0.023	0.022	0.96	0.017	0.017	0.98
	0.5	Ind	0.025	0.024	0.97	0.017	0.017	0.99	0.021	0.021	0.98	0.015	0.015	0.98
		Ar1	0.021	0.020	0.96	0.015	0.014	0.98	0.018	0.017	0.98	0.012	0.012	0.98
		Un	0.023	0.019	0.94	0.015	0.014	0.97	0.019	0.018	0.96	0.013	0.014	0.98
	0.9	Ind	0.025	0.024	0.97	0.016	0.017	0.98	0.023	0.022	0.95	0.016	0.015	0.98
		Ar1	0.010	0.009	0.98	0.007	0.006	0.99	0.009	0.008	0.98	0.006	0.006	1.00
		Un	0.016	0.015	0.87	0.008	0.007	0.96	0.014	0.011	0.91	0.008	0.007	0.98
0.50	0.1	Ind	0.035	0.032	0.97	0.026	0.023	0.95	0.030	0.029	0.96	0.022	0.021	0.96
		Ar1	0.034	0.032	0.96	0.026	0.023	0.95	0.030	0.029	0.95	0.022	0.021	0.96
		Un	0.035	0.030	0.93	0.026	0.023	0.95	0.031	0.029	0.95	0.022	0.023	0.97
	0.5	Ind	0.034	0.033	0.98	0.023	0.023	0.99	0.030	0.029	0.96	0.021	0.021	0.97
		Ar1	0.029	0.027	0.96	0.020	0.020	0.99	0.025	0.024	0.97	0.018	0.017	0.98
		Un	0.032	0.026	0.94	0.021	0.019	0.97	0.028	0.024	0.94	0.019	0.019	0.97
	0.9	Ind	0.034	0.032	0.98	0.022	0.023	0.98	0.030	0.029	0.96	0.022	0.020	0.97
		Ar1	0.013	0.013	0.98	0.009	0.009	0.99	0.012	0.011	0.98	0.008	0.008	0.99
		Un	0.024	0.018	0.85	0.014	0.010	0.91	0.019	0.042	0.89	0.014	0.021	0.94
0.75	0.1	Ind	0.050	0.045	0.96	0.037	0.033	0.93	0.041	0.041	0.95	0.031	0.030	0.95
		Ar1	0.049	0.045	0.96	0.037	0.033	0.93	0.041	0.040	0.95	0.031	0.030	0.96
		Un	0.050	0.041	0.93	0.037	0.031	0.93	0.043	0.040	0.94	0.031	0.032	0.96
	0.5	Ind	0.049	0.045	0.96	0.033	0.033	0.98	0.043	0.040	0.96	0.029	0.029	0.96
		Ar1	0.043	0.039	0.95	0.029	0.029	0.97	0.038	0.035	0.95	0.026	0.025	0.97
		Un	0.049	0.037	0.90	0.032	0.027	0.94	0.048	0.036	0.90	0.031	0.027	0.96
	0.9	Ind	0.050	0.044	0.95	0.031	0.032	0.96	0.042	0.039	0.95	0.032	0.029	0.96
		Ar1	0.020	0.019	0.96	0.014	0.014	0.98	0.018	0.017	0.96	0.013	0.012	0.98
		Un	0.043	0.060	0.79	0.028	0.018	0.82	0.037	0.035	0.84	0.025	0.030	0.90

Tableau 2.15: **Location-scale-shift scenario** – Standard deviation, asymptotic standard errors and probability coverage of the GEEE estimator with different correlation structures at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ and the error term $\varepsilon \sim t_3$.

			$m = 4$						$m \sim \mathcal{U}(3,7)$					
			50			100			50			100		
τ	ρ	GEEE	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$
0.25	0.1	Ind	0.047	0.036	0.91	0.033	0.028	0.92	0.042	0.031	0.89	0.032	0.026	0.95
		Ar1	0.047	0.037	0.91	0.033	0.028	0.92	0.042	0.031	0.89	0.032	0.026	0.96
		Un	0.048	0.027	0.78	0.033	0.022	0.86	0.043	0.021	0.76	0.032	0.020	0.84
	0.5	Ind	0.050	0.038	0.92	0.035	0.028	0.92	0.057	0.038	0.92	0.037	0.028	0.93
		Ar1	0.051	0.038	0.91	0.035	0.029	0.93	0.057	0.038	0.93	0.037	0.028	0.95
		Un	0.053	0.028	0.78	0.036	0.023	0.86	0.059	0.024	0.80	0.038	0.021	0.85
	0.9	Ind	0.056	0.041	0.91	0.035	0.030	0.92	0.052	0.038	0.91	0.038	0.032	0.95
		Ar1	0.056	0.040	0.90	0.035	0.030	0.91	0.053	0.038	0.91	0.038	0.031	0.94
		Un	0.062	0.030	0.79	0.038	0.024	0.84	0.059	0.028	0.76	0.042	0.025	0.81
0.50	0.1	Ind	0.041	0.033	0.94	0.029	0.025	0.96	0.035	0.028	0.94	0.026	0.023	0.97
		Ar1	0.041	0.033	0.94	0.030	0.025	0.95	0.036	0.028	0.94	0.027	0.023	0.97
		Un	0.042	0.025	0.84	0.030	0.021	0.88	0.036	0.020	0.79	0.027	0.017	0.88
	0.5	Ind	0.042	0.034	0.95	0.031	0.026	0.96	0.048	0.034	0.94	0.028	0.024	0.96
		Ar1	0.043	0.035	0.95	0.031	0.026	0.96	0.049	0.034	0.95	0.028	0.025	0.96
		Un	0.045	0.026	0.84	0.032	0.021	0.89	0.050	0.021	0.80	0.029	0.019	0.88
	0.9	Ind	0.045	0.036	0.95	0.031	0.027	0.96	0.040	0.034	0.94	0.030	0.027	0.97
		Ar1	0.044	0.035	0.94	0.030	0.027	0.96	0.040	0.033	0.92	0.030	0.026	0.97
		Un	0.048	0.027	0.81	0.033	0.022	0.87	0.043	0.028	0.76	0.030	0.020	0.87
0.75	0.1	Ind	0.054	0.037	0.90	0.038	0.030	0.92	0.041	0.032	0.90	0.034	0.026	0.92
		Ar1	0.054	0.037	0.90	0.039	0.030	0.91	0.042	0.032	0.91	0.034	0.026	0.91
		Un	0.055	0.027	0.80	0.039	0.023	0.85	0.043	0.027	0.73	0.035	0.020	0.84
	0.5	Ind	0.056	0.038	0.92	0.041	0.030	0.95	0.078	0.039	0.90	0.034	0.029	0.94
		Ar1	0.057	0.038	0.92	0.041	0.031	0.95	0.078	0.039	0.91	0.034	0.029	0.94
		Un	0.060	0.029	0.77	0.042	0.023	0.89	0.082	0.024	0.76	0.035	0.098	0.83
	0.9	Ind	0.051	0.039	0.90	0.040	0.032	0.92	0.046	0.037	0.90	0.034	0.030	0.93
		Ar1	0.049	0.038	0.90	0.037	0.031	0.92	0.047	0.036	0.88	0.033	0.029	0.93
		Un	0.053	0.034	0.74	0.042	0.025	0.83	0.048	0.056	0.72	0.033	0.022	0.83

Tableau 2.16: **Location-scale-shift scenario** – Standard deviation, asymptotic standard errors and probability coverage of the GEEE estimator with different correlation structures at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ and the error term $\varepsilon \sim \chi^2_3$.

			$m = 4$						$m \sim \mathcal{U}(3,7)$					
			50			100			50			100		
τ	ρ	GEEE	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$	SD	SE	$P_{0.95}$
0.25	0.1	Ind	0.043	0.037	0.91	0.030	0.029	0.97	0.039	0.034	0.90	0.029	0.026	0.96
		Ar1	0.044	0.037	0.90	0.030	0.029	0.96	0.039	0.034	0.90	0.030	0.026	0.95
		Un	0.044	0.031	0.84	0.031	0.026	0.92	0.041	0.025	0.80	0.030	0.023	0.92
	0.5	Ind	0.047	0.039	0.94	0.033	0.030	0.92	0.040	0.036	0.94	0.029	0.027	0.95
		Ar1	0.048	0.039	0.92	0.034	0.031	0.90	0.041	0.037	0.93	0.030	0.028	0.94
		Un	0.050	0.033	0.85	0.035	0.027	0.89	0.043	0.027	0.84	0.031	0.024	0.89
	0.9	Ind	0.051	0.042	0.90	0.034	0.032	0.95	0.044	0.041	0.94	0.031	0.030	0.96
		Ar1	0.052	0.042	0.90	0.034	0.032	0.96	0.044	0.041	0.92	0.032	0.030	0.96
		Un	0.054	0.035	0.82	0.035	0.028	0.91	0.046	0.032	0.81	0.033	0.026	0.90
0.50	0.1	Ind	0.058	0.049	0.90	0.042	0.039	0.96	0.054	0.045	0.88	0.040	0.035	0.95
		Ar1	0.059	0.049	0.90	0.042	0.039	0.96	0.055	0.045	0.87	0.040	0.036	0.95
		Un	0.059	0.040	0.84	0.043	0.033	0.91	0.056	0.033	0.75	0.041	0.029	0.92
	0.5	Ind	0.064	0.051	0.91	0.045	0.040	0.91	0.054	0.048	0.92	0.041	0.037	0.94
		Ar1	0.066	0.052	0.89	0.046	0.041	0.91	0.056	0.049	0.92	0.042	0.038	0.94
		Un	0.069	0.042	0.79	0.047	0.035	0.86	0.057	0.034	0.82	0.046	0.030	0.87
	0.9	Ind	0.069	0.056	0.88	0.046	0.042	0.93	0.061	0.054	0.91	0.043	0.040	0.95
		Ar1	0.070	0.056	0.89	0.047	0.042	0.92	0.062	0.053	0.89	0.043	0.040	0.94
		Un	0.073	0.047	0.77	0.049	0.036	0.85	0.064	0.046	0.76	0.045	0.036	0.89
0.75	0.1	Ind	0.079	0.065	0.87	0.060	0.052	0.92	0.076	0.058	0.84	0.057	0.048	0.92
		Ar1	0.079	0.065	0.86	0.061	0.053	0.92	0.077	0.058	0.84	0.057	0.048	0.94
		Un	0.081	0.051	0.80	0.061	0.042	0.86	0.077	0.046	0.72	0.057	0.039	0.85
	0.5	Ind	0.087	0.067	0.84	0.063	0.053	0.90	0.075	0.063	0.90	0.058	0.049	0.92
		Ar1	0.090	0.068	0.85	0.064	0.054	0.90	0.077	0.064	0.90	0.059	0.050	0.92
		Un	0.094	0.059	0.76	0.065	0.043	0.83	0.078	0.045	0.78	0.060	0.040	0.85
	0.9	Ind	0.096	0.073	0.82	0.065	0.056	0.90	0.085	0.071	0.86	0.060	0.054	0.90
		Ar1	0.096	0.071	0.82	0.065	0.056	0.89	0.086	0.069	0.86	0.061	0.053	0.91
		Un	0.102	0.061	0.71	0.070	0.064	0.80	0.091	0.094	0.74	0.065	0.055	0.84

Tableau 2.17: **Location-shift scenario** – Bias and probability coverage of the GEEE estimator with exchangeable working correlation and the LQMM estimator at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ for a **balanced** panel.

			$n = 50$				$n = 100$			
			Bias		$P_{0.95}$		Bias		$P_{0.95}$	
	ρ	τ	Exc	LQMM	Exc	LQMM	Exc	LQMM	Exc	LQMM
$\mathcal{N}$	0.1	τ_1	-0.0005	-0.0008	0.97	1.00	-0.0002	-0.0001	0.98	0.99
		τ_2	-0.0006	-0.0008	0.98	0.99	0.0000	0.0000	0.99	0.98
		τ_3	-0.0006	-0.0008	0.97	0.98	0.0001	0.0001	0.99	0.99
	0.5	τ_1	0.0005	0.0000	0.97	0.99	-0.0007	0.0000	0.98	0.99
		τ_2	0.0003	0.0007	0.97	0.98	-0.0005	-0.0005	0.99	0.99
		τ_3	0.0001	0.0000	0.98	0.98	-0.0004	-0.0002	0.99	0.99
	0.9	τ_1	0.0001	0.0002	0.99	0.99	0.0003	-0.0003	0.99	0.99
		τ_2	0.0001	0.0000	1.00	0.99	0.0004	0.0005	1.00	0.99
		τ_3	0.0002	-0.0001	0.98	0.98	0.0005	0.0008	0.99	1.00
$\mathcal{T}_3$	0.1	τ_1	0.0000	-0.0001	0.97	0.97	0.0021	-0.0005	0.98	0.96
		τ_2	0.0001	-0.0015	0.97	0.97	0.0015	-0.0013	0.98	0.97
		τ_3	0.0004	-0.0049	0.97	0.98	0.0016	-0.0020	0.97	0.97
	0.5	τ_1	-0.0010	-0.0010	0.98	0.98	-0.0005	0.0007	0.98	0.98
		τ_2	-0.0001	-0.0027	0.98	0.97	-0.0004	-0.0008	0.98	0.98
		τ_3	0.0003	-0.0026	0.97	0.98	-0.0001	0.0007	0.98	0.97
	0.9	τ_1	-0.0001	-0.0005	0.97	0.99	-0.0004	0.0004	0.98	0.99
		τ_2	0.0000	-0.0010	0.99	0.99	-0.0005	-0.0001	0.99	0.98
		τ_3	0.0002	-0.0010	0.99	0.97	-0.0007	0.0000	0.98	0.99
χ^3_2	0.1	τ_1	-0.0005	0.0376	0.98	0.62	-0.0003	0.0381	0.96	0.34
		τ_2	-0.0016	0.0325	0.97	0.71	-0.0005	0.0340	0.95	0.48
		τ_3	-0.0027	0.0284	0.96	0.83	-0.0008	0.0295	0.93	0.66
	0.5	τ_1	-0.0016	0.0352	0.97	0.70	-0.0003	0.0371	0.98	0.44
		τ_2	-0.0024	0.0294	0.96	0.82	-0.0001	0.0312	0.98	0.59
		τ_3	-0.0032	0.0238	0.96	0.91	0.0004	0.0251	0.97	0.79
	0.9	τ_1	-0.0006	0.0341	0.98	0.78	0.0000	0.0374	0.99	0.48
		τ_2	-0.0009	0.0212	0.98	0.88	0.0001	0.0237	0.98	0.76
		τ_3	-0.0016	0.0088	0.96	0.96	0.0004	0.0095	0.97	0.96

Tableau 2.18: **Location-shift scenario** – Bias and probability coverage of the GEEE estimator with exchangeable working correlation and the LQMM estimator at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ for an **unbalanced**.

			$n = 50$				$n = 100$			
			Bias		$P_{0.95}$		Bias		$P_{0.95}$	
	ρ	τ	Exc	LQMM	Exc	LQMM	Exc	LQMM	Exc	LQMM
$\mathcal{N}$	0.1	τ_1	0.0007	0.0010	0.97	0.97	-0.0006	-0.0006	0.99	0.99
		τ_2	0.0005	0.0009	0.97	0.99	-0.0008	-0.0009	0.99	1.00
		τ_3	0.0001	0.0001	0.97	0.97	-0.0009	-0.0009	0.98	1.00
	0.5	τ_1	-0.0002	-0.0003	0.98	0.99	0.0002	0.0001	1.00	0.99
		τ_2	-0.0003	-0.0006	0.99	0.99	0.0002	0.0001	1.00	0.99
		τ_3	-0.0004	0.0004	0.97	0.98	0.0002	0.0000	1.00	1.00
	0.9	τ_1	0.0005	0.0009	0.99	0.98	0.0000	0.0003	1.00	1.00
		τ_2	0.0003	0.0007	0.99	0.99	0.0000	-0.0006	1.00	1.00
		τ_3	0.0000	0.0003	0.99	1.00	0.0000	0.0003	1.00	1.00
$\mathcal{T}_3$	0.1	τ_1	-0.0015	-0.0014	0.97	0.98	0.0005	-0.0011	0.98	0.96
		τ_2	-0.0012	-0.0011	0.98	0.96	0.0004	-0.0008	0.99	0.96
		τ_3	-0.0010	-0.0020	0.97	0.97	-0.0001	-0.0013	0.98	0.97
	0.5	τ_1	0.0006	0.0001	0.98	0.99	0.0014	0.0003	0.98	0.98
		τ_2	0.0006	-0.0001	0.98	0.98	0.0006	0.0011	0.99	0.98
		τ_3	0.0009	0.0033	0.98	0.98	0.0000	0.0005	0.98	0.98
	0.9	τ_1	-0.0009	-0.0004	0.99	0.98	0.0001	0.0003	0.99	0.99
		τ_2	-0.0003	-0.0005	0.98	0.98	0.0004	0.0004	1.00	0.99
		τ_3	0.0002	0.0001	0.98	0.97	0.0008	0.0013	1.00	0.98
χ^3_2	0.1	τ_1	-0.0008	0.0383	0.96	0.56	-0.0012	0.0371	0.96	0.24
		τ_2	-0.0010	0.0336	0.96	0.68	-0.0014	0.0337	0.96	0.34
		τ_3	-0.0008	0.0299	0.96	0.74	-0.0018	0.0301	0.96	0.52
	0.5	τ_1	0.0007	0.0365	0.97	0.66	0.0010	0.0370	0.98	0.34
		τ_2	0.0011	0.0303	0.96	0.76	0.0012	0.0311	0.98	0.55
		τ_3	0.0012	0.0257	0.95	0.86	0.0013	0.0261	0.97	0.73
	0.9	τ_1	-0.0008	0.0355	0.97	0.70	0.0002	0.0361	0.99	0.47
		τ_2	-0.0006	0.0219	0.97	0.87	0.0003	0.0236	1.00	0.81
		τ_3	-0.0004	0.0082	0.95	0.96	0.0005	0.0120	0.98	0.97

Tableau 2.19: **Location-scale-shift scenario** – Bias and probability coverage of the GEEE estimator with exchangeable working correlation and the LQMM estimator at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ for a **balanced** panel.

			$n = 50$				$n = 100$			
			Bias		$P_{0.95}$		Bias		$P_{0.95}$	
	ρ	τ	Exc	LQMM	Exc	LQMM	Exc	LQMM	Exc	LQMM
$\mathcal{N}$	0.1	τ_1	-0.0008	-0.0003	0.99	0.99	-0.0007	0.0007	0.99	0.98
		τ_2	-0.0001	-0.0005	0.99	1.00	0.0000	0.0001	1.00	0.99
		τ_3	0.0006	0.0000	0.98	0.99	0.0008	0.0005	0.98	0.98
	0.5	τ_1	-0.0019	0.0008	0.98	0.99	-0.0018	0.0002	0.97	0.99
		τ_2	-0.0002	0.0008	1.00	0.98	-0.0001	-0.0002	1.00	0.98
		τ_3	0.0015	0.0006	0.98	0.99	0.0016	0.0002	0.98	0.98
	0.9	τ_1	-0.0018	0.0004	0.97	0.99	-0.0017	0.0000	0.99	1.00
		τ_2	-0.0001	-0.0003	0.98	0.99	0.0000	-0.0004	1.00	0.99
		τ_3	0.0017	-0.0005	0.98	0.99	0.0017	-0.0008	0.97	0.99
$\mathcal{T}_3$	0.1	τ_1	-0.0011	-0.0003	1.00	0.98	-0.0011	-0.0016	0.99	0.98
		τ_2	0.0001	0.0004	1.00	0.99	0.0000	-0.0010	0.99	0.97
		τ_3	0.0013	-0.0011	1.00	0.98	0.0012	-0.0010	1.00	0.98
	0.5	τ_1	-0.0025	0.0027	1.00	0.98	-0.0019	0.0023	1.00	0.98
		τ_2	0.0000	0.0017	1.00	0.98	0.0003	0.0011	1.00	0.98
		τ_3	0.0025	0.0007	1.00	0.97	0.0026	-0.0006	0.99	0.98
	0.9	τ_1	-0.0026	0.0155	0.98	0.96	-0.0025	0.0149	0.98	0.96
		τ_2	-0.0001	-0.0001	0.98	0.99	0.0000	0.0001	0.99	0.98
		τ_3	0.0025	-0.0150	0.99	0.96	0.0025	-0.0156	0.99	0.94
χ^3_2	0.1	τ_1	-0.0015	-0.0010	0.98	0.95	-0.0015	0.0033	0.98	0.98
		τ_2	-0.0002	-0.0027	0.98	0.97	0.0001	0.0059	1.00	0.99
		τ_3	0.0019	-0.0034	0.98	0.98	0.0020	0.0057	1.00	0.97
	0.5	τ_1	-0.0036	0.0015	0.96	0.96	-0.0034	0.0022	0.96	0.97
		τ_2	-0.0002	0.0001	0.98	0.96	0.0001	0.0003	1.00	0.96
		τ_3	0.0043	-0.0101	0.98	0.95	0.0044	-0.0094	0.99	0.96
	0.9	τ_1	-0.0037	0.0101	0.98	0.96	-0.0035	0.0149	0.96	0.95
		τ_2	-0.0001	-0.0112	0.99	0.97	0.0001	-0.0061	0.99	0.96
		τ_3	0.0045	-0.0433	0.98	0.92	0.0047	-0.0398	0.98	0.91

Tableau 2.20: **Location-scale-shift scenario** – Bias and probability coverage of the GEEE estimator with exchangeable working correlation and the LQMM estimator at 3 percentiles when $\rho \in (0.1, 0.5, 0.9)$ for an **unbalanced** panel.

			$n = 50$				$n = 100$			
			Bias		$P_{0.95}$		Bias		$P_{0.95}$	
	ρ	τ	Exc	LQMM	Exc	LQMM	Exc	LQMM	Exc	LQMM
$\mathcal{N}$	0.1	τ_1	-0.0007	-0.0015	0.99	0.99	-0.0008	0.0007	1.00	0.99
		τ_2	0.0000	-0.0014	0.99	0.99	-0.0001	0.0011	1.00	0.99
		τ_3	0.0006	-0.0007	0.98	0.98	0.0005	0.0009	1.00	1.00
	0.5	τ_1	-0.0015	-0.0006	0.99	0.98	-0.0015	0.0013	0.99	0.99
		τ_2	-0.0001	0.0001	1.00	0.98	0.0000	0.0008	1.00	0.98
		τ_3	0.0013	-0.0009	0.99	0.99	0.0014	-0.0003	0.99	0.99
	0.9	τ_1	-0.0017	0.0004	0.97	0.98	-0.0017	-0.0002	0.99	1.00
		τ_2	0.0000	-0.0005	0.98	0.98	0.0000	0.0004	1.00	1.00
		τ_3	0.0017	0.0008	0.97	0.97	0.0016	0.0002	0.98	0.99
$\mathcal{T}_3$	0.1	τ_1	-0.0009	0.0007	0.99	0.97	-0.0012	-0.0008	1.00	0.99
		τ_2	0.0000	-0.0002	0.99	0.96	-0.0002	-0.0003	1.00	0.97
		τ_3	0.0010	-0.0001	0.98	0.96	0.0007	0.0008	1.00	0.97
	0.5	τ_1	-0.0022	0.0019	1.00	0.97	-0.0017	0.0001	1.00	0.98
		τ_2	0.0001	0.0005	1.00	0.97	0.0002	-0.0005	0.99	0.99
		τ_3	0.0025	-0.0005	1.00	0.96	0.0022	-0.0008	1.00	0.98
	0.9	τ_1	-0.0028	0.0134	0.99	0.96	-0.0026	0.0161	0.99	0.93
		τ_2	-0.0003	-0.0013	0.98	0.98	-0.0001	0.0006	1.00	0.98
		τ_3	0.0020	-0.0172	0.99	0.94	0.0023	-0.0151	0.99	0.94
χ^3_2	0.1	τ_1	-0.0018	-0.0008	0.98	0.98	-0.0012	0.0030	0.98	0.98
		τ_2	-0.0005	-0.0027	0.99	0.96	0.0002	0.0022	1.00	0.96
		τ_3	0.0013	-0.0035	0.99	0.96	0.0022	0.0012	0.99	0.96
	0.5	τ_1	-0.0030	0.0020	0.98	0.97	-0.0028	0.0024	0.95	0.97
		τ_2	0.0000	-0.0017	0.99	0.97	0.0003	0.0007	0.99	0.98
		τ_3	0.0036	-0.0120	0.99	0.98	0.0039	-0.0087	0.98	0.96
	0.9	τ_1	-0.0035	0.0141	0.98	0.96	-0.0035	0.0121	0.92	0.96
		τ_2	0.0000	-0.0052	1.00	0.97	-0.0002	-0.0091	0.99	0.97
		τ_3	0.0046	-0.0337	0.98	0.93	0.0042	-0.0376	0.99	0.92

2.7.2 Supplementary Material II : Proof of the Theorems

We present the following **Lemma 2.7.0.1**, which is an important result for the proofs of different results in this thesis. It plays a central role in approximating our objective functions by quadratic functions. We give its proof right after the statement.

Lemma 2.7.0.1. *Let* $r(t) = \rho_\tau(u-t) - \rho_\tau(t) + 2t\psi_\tau(u)u$ *then*

$$r(t) = O(t^2).$$

Proof of Lemma 2.7.0.1.

Replacing the functions $\rho_\tau(\cdot)$, and $\psi_\tau(\cdot)$ by their expression, the function $r(\cdot)$ is

$$r(t) = |\tau - \mathbb{1}(u<t)|(u-t)^2 - |\tau - \mathbb{1}(u<0)|u^2 + 2|\tau - \mathbb{1}(u<0)|ut.$$

According to the sign of t, we can distinguish two cases. If $t > 0$, we have :

$$r(t) = \begin{cases} (1-\tau)t^2, & \text{if } u < 0 < t \\ (1-2\tau)(u-t)^2 + \tau t^2, & \text{if } 0 < u < t \\ \tau t^2. & \text{if } 0 < t < u \end{cases}$$

If $t < 0$, one has,

$$r(t) = \begin{cases} (1-\tau)t^2, & \text{if } u < t < 0 \\ (2\tau-1)(u-t)^2 + (1-\tau)t^2, & \text{if } t < u < 0 \\ \tau t^2. & \text{if } t < 0 < u \end{cases}$$

Then, we have $r(t) = O(t^2)$. □

Proof of Theorem 2.3.1.

Consider the following objective function

$$R_{Nq}(\boldsymbol{\delta}) = \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{t=1}^{m_i} w_k\left\{\rho_\tau\left(\varepsilon_{it\tau_k} - \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\delta}_{\tau_k}/\sqrt{N}\right) - \rho_\tau\left(\varepsilon_{it\tau_k}\right)\right\} \tag{2.22}$$

where $\boldsymbol{\delta} = (\boldsymbol{\delta}_{\tau_1}{}^{\mathsf{T}}, \ldots, \boldsymbol{\delta}_{\tau_q}{}^{\mathsf{T}})^{\mathsf{T}}$ is a $pq \times 1$ vector, $\boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}}$ a $p \times 1$ vector of parameters and $\varepsilon_{it\tau_k} = y_{it} - \mu_{it\tau_k}$ and $\mu_{it\tau_k} = \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}_{\tau_k}$.

R_{Nq} is a convex function of $\boldsymbol{\delta}$ and is minimized by $\widehat{\boldsymbol{\delta}}$

$$\widehat{\boldsymbol{\delta}} = \begin{pmatrix} \widehat{\boldsymbol{\delta}_{\tau_1}} \\ \vdots \\ \widehat{\boldsymbol{\delta}_{\tau_q}} \end{pmatrix} = \begin{pmatrix} \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\tau_1} - \boldsymbol{\beta}_{\tau_1}\right) \\ \vdots \\ \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\tau_q} - \boldsymbol{\beta}_{\tau_q}\right) \end{pmatrix}.$$

The idea is to approximate R_{Nq} by a convex quadratic function with a unique minimum and use result from Lid Hjort et Pollard (2011) to show that $\widehat{\boldsymbol{\delta}}$ has the same asymptotic distribution that minimizing value.

The convex function $\mathbb{E}\left[\rho_\tau\left(\varepsilon_{it\tau_k} - \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\delta}_{\tau_k}/\sqrt{N}\right) - \rho_\tau(\varepsilon_{it\tau_k})\right]$ is twice continuously differentiable and it reaches its minimum at $\boldsymbol{\delta}_{\tau_k} = \mathbf{0}$. It can be represented in the neighbourhood of $\boldsymbol{\delta}_{\tau_k} = \mathbf{0}$ as

$$\begin{aligned}\mathbb{E}\left[\rho_\tau\left(\varepsilon_{it\tau_k} - \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\delta}_{\tau_k}/\sqrt{N}\right) - \rho_\tau(\varepsilon_{it\tau_k})\right] &= -2N^{-1/2}\boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}}\boldsymbol{x}_{it}\,\mathbb{E}[\psi_\tau(\varepsilon_{it\tau_k}).\varepsilon_{it\tau_k}] \\ &\quad + N^{-1}\boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}}\boldsymbol{x}_{it}\,\mathbb{E}[\psi_\tau(\varepsilon_{it\tau_k})]\boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\delta}_{\tau_k} + o\left\|\boldsymbol{\delta}_{\tau_k}/\sqrt{N}\right\|^2,\end{aligned} \tag{2.23}$$

where $\psi_\tau(\lambda) = \tau - \mathbb{1}(\lambda < 0)$. Since

$$\operatorname{argmin}\mathbb{E}\left[\rho_\tau\left(\varepsilon_{it\tau_k} - \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\delta}_{\tau_k}/\sqrt{N}\right) - \rho_\tau(\varepsilon_{it\tau_k})\right] = \mathbf{0}$$

then by the first order condition

$$\mathbb{E}[\psi_\tau(\varepsilon_{it\tau_k})\varepsilon_{it\tau_k}] = 0, \tag{2.24}$$

and equation (2.23) can be reduced to

$$\mathbb{E}\left[\rho_\tau\left(\varepsilon_{it\tau_k} - \boldsymbol{x}_{it}^{\top}\boldsymbol{\delta}_{\tau_k}/\sqrt{N}\right) - \rho_\tau(\varepsilon_{it\tau_k})\right] = N^{-1}\boldsymbol{\delta}_{\tau_k}^{\top}\boldsymbol{x}_{it}\,\mathbb{E}[\psi_\tau(\varepsilon_{it\tau_k})]\boldsymbol{x}_{it}^{\top}\boldsymbol{\delta}_{\tau_k} + o\left\|\boldsymbol{\delta}_{\tau_k}/\sqrt{N}\right\|^2. \tag{2.25}$$

Define

$$D_{it}(\varepsilon_{it\tau_k}) = -2\psi_\tau(\varepsilon_{it\tau_k})\varepsilon_{it\tau_k}. \tag{2.26}$$

Notice that by (2.24), $\mathbb{E}[D_{it}(\varepsilon_{it\tau_k})] = 0$. Let

$$r_{it}(\boldsymbol{\delta}_{\tau_k}) = \rho_\tau\left(\varepsilon_{it\tau_k} - \boldsymbol{x}_{it}^{\top}\boldsymbol{\delta}_{\tau_k}/\sqrt{N}\right) - \rho_\tau(\varepsilon_{it\tau_k}) - \boldsymbol{\delta}_{\tau_k}^{\top}\boldsymbol{x}_{it}D_{it}(\varepsilon_{it\tau_k})/\sqrt{N}$$

then R_{Nq} can be written as

$$\begin{aligned} R_{Nq}(\boldsymbol{\delta}) = &\sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{t=1}^{m_i} w_k\boldsymbol{\delta}_{\tau_k}^{\top}\boldsymbol{x}_{it}D_{it}(\varepsilon_{it\tau_k})/\sqrt{N} \\ &+ \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{t=1}^{m_i} w_k\left(\mathbb{E}\left[\rho_\tau\left(\varepsilon_{it\tau_k} - \boldsymbol{x}_{it}^{\top}\boldsymbol{\delta}_{\tau_k}/\sqrt{N}\right) - \rho_\tau(\varepsilon_{it\tau_k})\right]\right) \\ &+ \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{t=1}^{m_i} w_k\left(r_{it}(\boldsymbol{\delta}_{\tau_k}) - \mathbb{E}\left[r_{it}(\boldsymbol{\delta}_{\tau_k})\right]\right). \end{aligned}$$

Using **Lemma 2.7.0.1** it follows that

$$
\begin{aligned}
R_{Nq}(\boldsymbol{\delta}) &= \frac{1}{\sqrt{N}} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{m_i} w_k \boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}} \boldsymbol{x}_{it} D_{it}(\varepsilon_{it\tau_k}) \\
&+ \frac{1}{N} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{m_i} w_k \boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}} \boldsymbol{x}_{it}\, \mathbb{E}[\psi_\tau(\varepsilon_{it\tau_k})] \boldsymbol{x}_{it}{}^{\mathsf{T}} \boldsymbol{\delta}_{\tau_k} + o\|\boldsymbol{\delta}_{\tau_k}\|^2 + O\|\boldsymbol{\delta}_{\tau_k}\|^2 \\
&= \frac{-2}{\sqrt{N}} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{m_i} w_k \boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}} \boldsymbol{x}_{it} \psi_\tau(\varepsilon_{it\tau_k}).\varepsilon_{it\tau_k} \\
&+ \frac{1}{N} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{m_i} w_k \boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}} \boldsymbol{x}_{it}\, \mathbb{E}[\psi_\tau(\varepsilon_{it\tau_k})] \boldsymbol{x}_{it}{}^{\mathsf{T}} \boldsymbol{\delta}_{\tau_k} + o_p(1) \\
&\simeq -2 \frac{1}{\sqrt{N}} \sum_{k=1}^{q} w_k \boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}} \boldsymbol{X}^{\mathsf{T}} \boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{\tau_k}) \boldsymbol{\varepsilon}_{\tau_k} + \frac{1}{N} \sum_{k=1}^{q} w_k \boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}} \boldsymbol{X}^{\mathsf{T}} \mathbb{E}[\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{\tau_k})] \boldsymbol{X} \boldsymbol{\delta}_{\tau_k} \\
&= R_{Nq}^{(1)}(\boldsymbol{\delta}) + R_{Nq}^{(2)}(\boldsymbol{\delta}).
\end{aligned}
$$

Now consider $R_{Nq}^{(1)}$ the first term to the right of R_{Nq}. We have

$$
\begin{aligned}
R_{Nq}^{(1)}(\boldsymbol{\delta}) &= \frac{-2}{\sqrt{N}}\boldsymbol{\delta}^{\mathsf{T}}(\boldsymbol{W}\otimes\boldsymbol{X})^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{\boldsymbol{\tau}})\boldsymbol{\varepsilon}_{\boldsymbol{\tau}} \\
&= \frac{-2}{\sqrt{N}}\boldsymbol{\delta}^{\mathsf{T}}\begin{pmatrix} w_1\boldsymbol{X}^{\mathsf{T}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & w_q\boldsymbol{X}^{\mathsf{T}} \end{pmatrix}\begin{pmatrix} \boldsymbol{\Psi}_{\tau_1}(\boldsymbol{\varepsilon}_{\tau_1}) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Psi}_{\tau_q}(\boldsymbol{\varepsilon}_{\tau_q}) \end{pmatrix}\begin{pmatrix} \boldsymbol{\varepsilon}_{\tau_1} \\ \vdots \\ \boldsymbol{\varepsilon}_{\tau_q} \end{pmatrix} \\
&= \frac{-2}{\sqrt{N}}\boldsymbol{\delta}^{\mathsf{T}}\sum_{i=1}^{n}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}})\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}} \\
&= \frac{-2}{\sqrt{N}}\boldsymbol{\delta}^{\mathsf{T}}\sum_{i=1}^{n}\begin{bmatrix} w_1\sum_{t=1}^{m_i}x_{it}^1\psi_{\tau_1}(\varepsilon_{it\tau_1})\varepsilon_{it\tau_1} \\ \vdots \\ w_1\sum_{t=1}^{m_i}x_{it}^p\psi_{\tau_1}(\varepsilon_{it\tau_1})\varepsilon_{it\tau_1} \\ \vdots \\ w_q\sum_{t=1}^{m_i}x_{it}^1\psi_{\tau_q}(\varepsilon_{it\tau_q})\varepsilon_{it\tau_q} \\ \vdots \\ w_q\sum_{t=1}^{m_i}x_{it}^p\psi_{\tau_q}(\varepsilon_{it\tau_q})\varepsilon_{it\tau_q} \end{bmatrix} \\
&\xrightarrow{d} -2\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{B}
\end{aligned}
$$

To prove the asymptotic normality of $\boldsymbol{B}$, we apply the Cramér-Wold device, **Theorem A.2**, and verify the Lyapunov's condition, **Theorem A.3**.

Let $Z_{ni} = \boldsymbol{\lambda}^{\mathsf{T}}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}})\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}$ and consider $n^{-1/2}\sum_{i=1}^{n}Z_{ni}$, where $\boldsymbol{\lambda}$ is a $pq\times 1$ unit vector, $\boldsymbol{\lambda}^{\mathsf{T}}\boldsymbol{\lambda} = 1$. The summands Z_{ni} are independent with $\mathbb{E}[Z_{ni}] = 0$ and $\mathrm{Var}\left[n^{-1/2}\sum_{i=1}^{n}Z_{ni}\right] > \nu\prime > 0$, by condition **A2**. By the Minkowski's inequality, we

have

$$\begin{aligned}\mathbb{E}|Z_{ni}|^{2+\nu} &= \mathbb{E}\left|\sum_{k=1}^{q}\sum_{l=1}^{p} w_k\lambda_{kl}\sum_{t=1}^{m_i} x_{it}^l\psi_{\tau_k}(\varepsilon_{it\tau_k})\varepsilon_{it\tau_k}\right|^{2+\nu} \\ &= \mathbb{E}\left|\sum_{k=1}^{q}\sum_{l=1}^{p}\sum_{t=1}^{m_i} w_k\lambda_{kl}x_{it}^l\psi_{\tau_k}(\varepsilon_{it\tau_k})\varepsilon_{it\tau_k}\right|^{2+\nu} \\ &\leq \left[\sum_{k=1}^{q}\sum_{l=1}^{p}\sum_{t=1}^{m_i}\left(\mathbb{E}\left|w_k\lambda_{kl}x_{it}^l\psi_{\tau_k}(\varepsilon_{it\tau_k})\varepsilon_{it\tau_k}\right|^{2+\nu}\right)^{\frac{1}{2+\nu}}\right]^{2+\nu}.\end{aligned}$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}\left|w_k\lambda_{kl}x_{it}^l\psi_{\tau_k}(\varepsilon_{it\tau_k})\varepsilon_{it\tau_k}\right|^{2+\nu} &= |w_k\lambda_{kl}x_{it}^l|^{2+\nu}\,\mathbb{E}\left|\psi_{\tau_k}(\varepsilon_{it\tau_k})\varepsilon_{it\tau_k}\right|^{2+\nu} \\ &\leq (M|\lambda_{kl}|)^{2+\nu}\left[\mathbb{E}|\psi_{\tau_k}(\varepsilon_{it\tau_k})|^{4+\nu}\right]^{1/2}\left[\mathbb{E}|\varepsilon_{it\tau_k}|^{4+\nu}\right]^{1/2} \\ &\leq (M|\lambda_{kl}|)^{2+\nu}\Delta,\end{aligned}$$

where the last inequality follows by $\mathbb{E}|\psi_{\tau_k}(\varepsilon_{it\tau_k})|^{4+\nu} < \Delta$ and $\mathbb{E}|\varepsilon_{it\tau_k}|^{4+\nu} < \Delta$. Hence,

$$\begin{aligned}\mathbb{E}|Z_{ni}|^{2+\nu} &\leq \left[\sum_{k=1}^{q}\sum_{l=1}^{p}\sum_{t=1}^{m_i} M|\lambda_{kl}|\Delta^{\frac{1}{2+\nu}}\right]^{2+\nu} \\ &\leq \Delta(Mm)^{2+\nu}|\boldsymbol{\lambda}^{\mathsf{T}}\mathbb{1}_{pq}|^{2+\nu} \\ &\leq \Delta m^{2+\nu}(pq)^{1+\nu}.\end{aligned}$$

Then by the Liapounov CLT, **Theorem A.3**, $\boldsymbol{B}$ is a zero mean Gaussian vector with covariance matrix $\boldsymbol{D}_{I0}(\boldsymbol{\tau})$. For the second term to the right of R_{Nq} we have, by condition **A2**,

$$\begin{aligned}R_{Nq}^{(2)}(\boldsymbol{\delta}) &= N^{-1}\boldsymbol{\delta}^{\mathsf{T}}\sum_{i=1}^{n}(\boldsymbol{W}\otimes\boldsymbol{X}_i)^{\mathsf{T}}\,\mathbb{E}\left[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}})\right](\boldsymbol{I}_q\otimes\boldsymbol{X}_i)\boldsymbol{\delta} \\ &\to \boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{D}_{I1}(\boldsymbol{\tau})\boldsymbol{\delta}.\end{aligned}$$

Thus, the limiting form of the objective function is

$$R_{0q}(\boldsymbol{\delta}) = -2\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{B} + \boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{D}_{I1}(\boldsymbol{\tau})\boldsymbol{\delta}$$

where $\boldsymbol{B}$ is a zero mean Gaussian vector with covariance matrix $\boldsymbol{D}_{I0}(\boldsymbol{\tau})$. Application of **Corollary A.1** or **Theorem A.1** gives the result of **Theorem 2.3.1**. □

Proof of Theorem 2.3.2.

For the proof of **Theorem 2.3.2**, we must show the convergence of $\widehat{\boldsymbol{D}}_{I1}(\boldsymbol{\tau})$ and $\widehat{\boldsymbol{D}}_{I0}(\boldsymbol{\tau})$. $\widehat{\boldsymbol{D}}_{I1}(\boldsymbol{\tau})$ is a diagonal matrix of general term $N^{-1}w_k\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k})\boldsymbol{X}_i, 1<k<q$. It suffices to show the convergence of $N^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k})\boldsymbol{X}_i$ to obtain that of $\widehat{\boldsymbol{D}}_{I1}(\boldsymbol{\tau})$. We have

$$
\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k})\boldsymbol{X}_i-\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{i\tau_k})]\boldsymbol{X}_i\right|\\
&\quad\leq\left|\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}[\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k})-\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{i\tau_k})]\boldsymbol{X}_i\right|\\
&\quad+\left|\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{i\tau_k})\boldsymbol{X}_i-\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{i\tau_k})]\boldsymbol{X}_i\right|\\
&\quad\leq\frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{m_i}\|\boldsymbol{x}_{it}\|^2\left|\psi_{\tau_k}(\widehat{\varepsilon}_{it\tau_k})-\psi_{\tau_k}(\varepsilon_{it\tau_k})\right| \qquad (2.27)\\
&\quad+\left|\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{i\tau_k})\boldsymbol{X}_i-\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{i\tau_k})]\boldsymbol{X}_i\right|\\
&\quad\leq\frac{M^2}{N}\sum_{i=1}^{n}\sum_{t=1}^{m_i}\mathbb{1}\Big(|\varepsilon_{it}|\leq pM|\widehat{\boldsymbol{\beta}}_{I\tau}-\boldsymbol{\beta}_{\tau}|\Big)\\
&\quad+\left|\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{i\tau_k})\boldsymbol{X}_i-\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{i\tau_k})]\boldsymbol{X}_i\right|.
\end{aligned}
$$

Since plim $\widehat{\boldsymbol{\beta}}_{I\tau}=\boldsymbol{\beta}_{\tau}$, then by the Markov's inequality the first term of the last equation of (2.27) converges in probability to 0 and by the Markov's Law of Large Numbers (LLN) the second term converges in probability to 0. Hence, $\widehat{\boldsymbol{D}}_{I1}(\boldsymbol{\tau})\xrightarrow{d}\boldsymbol{D}_{I1}(\boldsymbol{\tau})$.

The matrix $\widehat{\boldsymbol{D}}_{I0}(\boldsymbol{\tau})$ is a block matrix of dimension $pq \times pq$ and of general term $w_k w_j N^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \widehat{\boldsymbol{\Sigma}}_{i\tau_k\tau_j} \boldsymbol{X}_i$, $1 < k, j < q$, where $\widehat{\boldsymbol{\Sigma}}_{i\tau_k\tau_j} = \boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k}) \widehat{\boldsymbol{\varepsilon}}_{i\tau_k} \widehat{\boldsymbol{\varepsilon}}_{i\tau_j}{}^{\mathsf{T}} \boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_j})$. To obtain the convergence of $\widehat{\boldsymbol{D}}_{I0}(\boldsymbol{\tau})$, it suffices to show the convergence of its general term. To lighten the notation, we will show the convergence of $N^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \widehat{\boldsymbol{\Sigma}}_{i\tau} \boldsymbol{X}_i$. Without loss of generality, the case for $\tau_j \neq \tau_k$, that is $N^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \widehat{\boldsymbol{\Sigma}}_{i\tau_k\tau_j} \boldsymbol{X}_i$, is derived similarly.

Notice that $\widehat{\boldsymbol{\varepsilon}}_{i\tau} = \boldsymbol{\varepsilon}_{i\tau} - \boldsymbol{X}_i(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)$ and

$$\widehat{\boldsymbol{\varepsilon}}_{i\tau} \widehat{\boldsymbol{\varepsilon}}_{i\tau}{}^{\mathsf{T}} = \boldsymbol{\varepsilon}_{i\tau} \boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}} - \boldsymbol{\varepsilon}_{i\tau}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)^{\mathsf{T}} \boldsymbol{X}_i{}^{\mathsf{T}} - \boldsymbol{X}_i(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau) \boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}} + \boldsymbol{X}_i(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)^{\mathsf{T}} \boldsymbol{X}_i{}^{\mathsf{T}}.$$

With the above relationship let's break down the expression of $\frac{1}{N} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \widehat{\boldsymbol{\Sigma}}_{i\tau} \boldsymbol{X}_i$ as follows :

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \widehat{\boldsymbol{\Sigma}}_{i\tau} \boldsymbol{X}_i = {} & \frac{1}{N} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) \boldsymbol{\varepsilon}_{i\tau} \boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) \boldsymbol{X}_i \\
& - \frac{1}{N} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) \boldsymbol{\varepsilon}_{i\tau}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)^{\mathsf{T}} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) \boldsymbol{X}_i \\
& - \frac{1}{N} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) \boldsymbol{X}_i(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau) \boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) \boldsymbol{X}_i \\
& + \frac{1}{N} \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) \boldsymbol{X}_i(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)^{\mathsf{T}} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) \boldsymbol{X}_i.
\end{aligned}$$

Now let's develop the first term to the right of the previous equation as follows :

$$\begin{aligned}
& \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) \boldsymbol{\varepsilon}_{i\tau} \boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) \boldsymbol{X}_i = \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}) \boldsymbol{\varepsilon}_{i\tau} \boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}) \boldsymbol{X}_i \\
& + \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}) \boldsymbol{\varepsilon}_{i\tau} \boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}} (\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) - \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})) \boldsymbol{X}_i \\
& + \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} (\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) - \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})) \boldsymbol{\varepsilon}_{i\tau} \boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}} \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}) \boldsymbol{X}_i \\
& + \sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}} (\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) - \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})) \boldsymbol{\varepsilon}_{i\tau} \boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}} (\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) - \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})) \boldsymbol{X}_i.
\end{aligned}$$

The new expression of $\frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}_{i\tau}\boldsymbol{X}_i$ is :

$$\frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}_{i\tau}\boldsymbol{X}_i = \frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{X}_i \tag{e1}$$

$$- \frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)^{\mathsf{T}}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i \tag{e2}$$

$$- \frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)\boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i \tag{e3}$$

$$+ \frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)^{\mathsf{T}}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i \tag{e4}$$

$$+ \frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}}(\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) - \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}))\boldsymbol{X}_i \tag{e5}$$

$$+ \frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}}(\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) - \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}))\boldsymbol{\varepsilon}_{i\tau}\boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{X}_i \tag{e6}$$

$$+ \frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^{\mathsf{T}}(\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) - \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}))\boldsymbol{\varepsilon}_{i\tau}\boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}}(\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau}) - \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}))\boldsymbol{X}_i. \tag{e7}$$

The demonstration will consist of showing the convergence of each of these terms : (e2)-(e7). Note that to show the convergence of (e4) is to show that of (e2) and that of (e6) is equivalent to that of (e5) and (e7). In the following, we will show respectively the convergence of (e3), (e4) and (e6).

Using the relation, $\mathrm{Vec}(ABC) = (C^{\mathsf{T}} \otimes A)\,\mathrm{Vec}(B)$, equation (e3) is transformed

as follows :

$$\mathrm{Vec}\Big(\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i(\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{\beta}_\tau)\boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i\Big)$$
$$=\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\otimes\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i\,\mathrm{Vec}(\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{\beta}_\tau).$$

Applying **Lemma A.1**, we have

$$\mathbb{E}\left\|\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\otimes\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i\right\|^{1+\nu}$$
$$\leq\Big(\mathbb{E}\left\|\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\right\|^{2+2\nu}\mathbb{E}\left\|\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i\right\|^{2+2\nu}\Big)^{1/2}.$$

Repeated application of Minkowski's and Holder's inequalities shows that

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{\varepsilon}_i\right\|^{2+2\nu}&=\mathbb{E}\left[\sum_{k=1}^{p}\Big|\sum_{t=1}^{m_i}x_{it}^k\psi_\tau(\widehat{\varepsilon}_{it})\varepsilon_{it}\Big|^2\right]^{1+\nu}\\&\leq\left[\sum_{k=1}^{p}\left(\mathbb{E}\Big|\sum_{t=1}^{m_i}x_{it}^k\psi_\tau(\widehat{\varepsilon}_{it})\varepsilon_{it}\Big|^{2+2\nu}\right)^{\frac{1}{1+\nu}}\right]^{1+\nu}\\&\leq\left\{\sum_{k=1}^{p}\left[\sum_{t=1}^{m_i}\left(\mathbb{E}|x_{it}^k\psi_\tau(\widehat{\varepsilon}_{it})\varepsilon_{it}|^{2+2\nu}\right)^{\frac{1}{2+2\nu}}\right]^{2+2\nu}\right\}^{1+\nu}\\&\leq(p\Delta)^{1+\nu}(Mm)^{2+\nu}.\end{aligned}$$

The last inequality is obtained by applying the assumptions $\mathbb{E}|\psi_\tau(\widehat{\epsilon}_{it\tau})|^{4+\nu}<\Delta$ and $\mathbb{E}|\varepsilon_{it}|^{4+\nu}<\Delta$. Similarly

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i\right\|^{2+2\nu}&\leq\mathbb{E}\left\|\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau^{1/2}(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\right\|^{4+4\nu}\\&\leq\left[\sum_{k=1}^{p}\sum_{t=1}^{m_i}\left(\mathbb{E}|(\psi_\tau^{1/2}(\widehat{\epsilon}_{it\tau})x_{it}^k)^2|^{2+2\nu}\right)^{\frac{1}{2+2\nu}}\right]^{2+2\nu}\\&\leq(pm)^{2+\nu}\Delta.\end{aligned}$$

Then, by the Markov's LLN it follows that

$$\mathrm{Vec}\Big(\frac{1}{N}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i(\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{\beta}_\tau)\boldsymbol{\varepsilon}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i\Big)=\frac{1}{\sqrt{N}}O_p(1)O_p(1).$$

Considering that $\text{Vec}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau) = O_p(\sqrt{N})$ and that $\mathbb{E}\left\|\boldsymbol{X}_i{}^\top \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i\right\|^{2+2\nu} < (pm)^{2+\nu}\Delta$, term (e4) is

$$\begin{aligned}
&\frac{1}{N}\text{Vec}\Big(\sum_{i=1}^{n} \boldsymbol{X}_i{}^\top \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)^\top \boldsymbol{X}_i{}^\top \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i\Big) \\
&= \frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^\top \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i \otimes \boldsymbol{X}_i{}^\top \boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_{i\tau})\boldsymbol{X}_i \,\text{Vec}[(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)^\top] \\
&= \frac{1}{N} O_p(1)O_p(1).
\end{aligned}$$

In the same way (e2) converge to 0. Now consider (e6)

$$\begin{aligned}
\left\|\boldsymbol{x}_{it}\boldsymbol{x}_{ik}{}^\top(\psi_\tau(\widehat{\varepsilon}_{it}) - \psi_\tau(\varepsilon_{it}))\psi_\tau(\varepsilon_{ik})\varepsilon_{it}\varepsilon_{ik}\right\| &\leq \|\boldsymbol{x}_{it}\|^2 |\varepsilon_{it}|^2 \mathbb{1}\Big(|\varepsilon_{it}| \leq p\|\boldsymbol{x}_{it}\| \left\|\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau\right\|\Big) \\
&\leq M^4 \mathbb{1}\Big(|\varepsilon_{it}| \leq pM \left\|\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau\right\|\Big)
\end{aligned}$$

where M is a generic constant. Since plim $\widehat{\boldsymbol{\beta}}_\tau = \boldsymbol{\beta}_\tau$, as shown above by the Markov's inequality, the expectation of (e6) converge to 0. In the same way (e5) and (e7) converge to 0.

We have just shown that

$$\frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^\top \widehat{\boldsymbol{\Sigma}}_{i\tau}\boldsymbol{X}_i - \frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^\top \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\boldsymbol{\varepsilon}_{i\tau}{}^\top \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{X}_i \xrightarrow{p} \mathbf{0}.$$

Application of the Markov's LLN also yields

$$\frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^\top \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\boldsymbol{\varepsilon}_{i\tau}{}^\top \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{X}_i - \frac{1}{N}\sum_{i=1}^{n} \boldsymbol{X}_i{}^\top \boldsymbol{\Sigma}_{i\tau}\boldsymbol{X}_i \xrightarrow{p} \mathbf{0}.$$

Thus, application of the triangular inequality shows that $\widehat{\boldsymbol{D}}_{I0}(\boldsymbol{\tau}) \xrightarrow{p} \boldsymbol{D}_{I0}(\boldsymbol{\tau})$. □

Proof of Theorem 2.3.3.

We will next present the proof of **Theorem 2.3.3**. We will start by showing the convergence in probability of $\boldsymbol{S}$

$$\boldsymbol{S}(\boldsymbol{\beta_\tau}) = \sum_{i=1}^{n} (\boldsymbol{W} \otimes \boldsymbol{X}_i)^\top \boldsymbol{V}_{i\tau}^{-1} \boldsymbol{\Psi_\tau} \Big(\mathbb{1}_q \otimes \boldsymbol{y}_i - (\mathbb{I}_q \otimes \boldsymbol{X}_i)\boldsymbol{\beta_\tau} \Big) \Big[\mathbb{1}_q \otimes \boldsymbol{y}_i - (\mathbb{I}_q \otimes \boldsymbol{X}_i)\boldsymbol{\beta_\tau} \Big],$$

which is a block diagonal matrix. Its convergence will be derived element-wise. Let $\boldsymbol{S}(\boldsymbol{\beta}_\tau) = \sum_{i=1}^{n} \boldsymbol{X}_i{}^\top \boldsymbol{V}_{i\tau}^{-1} \boldsymbol{\Psi}_\tau (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}_\tau)\big[\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}_\tau\big] = \sum_{i=1}^{n} \boldsymbol{S}_i(\boldsymbol{\beta}_\tau)$ be the estimating function for a single τ and denote the elements of the positive definite matrix as $\boldsymbol{V}_{i\tau}^{-1} = (\sigma_{its})_{1<t,s<m_i}$.

Repeated application of Minkowski's inequality and application of assumptions $\mathbb{E}|\psi_\tau(\varepsilon_{it\tau})|^{4+\nu} < \Delta$ and $\mathbb{E}|\varepsilon_{it\tau}|^{4+\nu} < \Delta$ show that

$$\begin{aligned}
\mathbb{E}\,\|\boldsymbol{S}_i\|^{2+2\nu} &= \mathbb{E}\left\{ \sum_{j=1}^{p} \left[\sum_{t=1}^{m_i} \sum_{s=1}^{m_i} \left| x_{it}^j \sigma_{its} \psi_\tau(\varepsilon_{is\tau}) \varepsilon_{is\tau} \right| \right]^2 \right\}^{1+\nu} \\
&\leq \left\{ \sum_{j=1}^{p} \left[\mathbb{E} \left(\sum_{t=1}^{m_i} \sum_{s=1}^{m_i} \left| x_{it}^j \sigma_{its} \psi_\tau(\varepsilon_{is\tau}) \varepsilon_{is\tau} \right| \right)^{2+2\nu} \right]^{\frac{1}{1+\nu}} \right\}^{1+\nu} \\
&\leq \left\{ \sum_{j=1}^{p} \left[\sum_{t=1}^{m_i} \sum_{s=1}^{m_i} \left(\mathbb{E} \left| x_{it}^j \sigma_{its} \psi_\tau(\varepsilon_{is\tau}) \varepsilon_{is\tau} \right|^{2+2\nu} \right)^{\frac{1}{2+2\nu}} \right]^2 \right\}^{1+\nu} \\
&\leq \left\{ \sum_{j=1}^{p} \left[M \Delta^{\frac{1}{2+2\nu}} \sum_{t=1}^{m_i} \sum_{s=1}^{m_i} |\sigma_{its}| \right]^2 \right\}^{1+\nu} \\
&\leq (pM^2)^{1+\nu} \Delta \Big(\sum_{t=1}^{m_i} \sum_{s=1}^{m_i} |\sigma_{its}| \Big)^{2+2\nu} \\
&\leq (pM^2)^{1+\nu} \Delta \left\| \boldsymbol{V}_{i\tau}^{-1} \right\|_1^{2+2\nu} \\
&\leq (pmM^2)^{1+\nu} \Delta
\end{aligned}$$

where the last inequality is derived from the equivalence of norms. This gives as the Markov's LLN condition

$$\mathbb{E} \left\| \boldsymbol{X}_i{}^\top \boldsymbol{V}_{i\tau}^{-1} \boldsymbol{\Psi}_\tau (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}_\tau)\big[\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}_\tau\big] \right\|^{2+2\delta} < \infty.$$

Therefore, by the Markov's LLN $N^{-1}\{\boldsymbol{S}(\boldsymbol{\beta}_\tau) - \mathbb{E}[\boldsymbol{S}(\boldsymbol{\beta}_\tau)]\} \xrightarrow{p} o_p(1)$. Since $\boldsymbol{S}(\boldsymbol{\beta}_\tau)$ is a convex function of $\boldsymbol{\beta}_\tau$ the convergence is uniform by the convexity **Lemma** of Pollard (1991a).

Furthermore $\mathbb{E}[\boldsymbol{S}(\boldsymbol{\beta}_\tau)]$ is continuously differentiable with

$$\boldsymbol{D}_{1\tau}(\boldsymbol{\beta}_\tau) = \frac{\partial \mathbb{E}[\boldsymbol{S}(\boldsymbol{\beta}_\tau)]}{\partial \boldsymbol{\beta}_\tau} = -\sum_{i=1}^{n} \boldsymbol{X}_i^\top \boldsymbol{V}_{i\tau}^{-1} \mathbb{E}\left[\boldsymbol{\Psi}_\tau(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_\tau)\right]\boldsymbol{X}_i$$

which is negative definite, under condition **B2**. Since $\mathbb{E}\left[\psi_\tau(y_{it} - \boldsymbol{x}_{it}\boldsymbol{\beta}_\tau)[(y_{it} - \boldsymbol{x}_{it}\boldsymbol{\beta}_\tau)]\right] = 0$, then $\boldsymbol{\beta}_\tau$ is the unique solution $\mathbb{E}[\boldsymbol{S}(\boldsymbol{\beta})] = 0$.

In summary, we have shown that the convex function $\boldsymbol{S}(\boldsymbol{\beta}_\tau)$ converges in probability uniformly to $\mathbb{E}[\boldsymbol{S}(\boldsymbol{\beta}_\tau)]$ which has a unique minimum then, due to to **Lemma A** of Newey et Powell (1987), we have $\widehat{\boldsymbol{\beta}}_\tau$, the solution of $\boldsymbol{S}(\boldsymbol{\beta}) = 0$ converges in probability to $\boldsymbol{\beta}_\tau$ unique solution of $\mathbb{E}[\boldsymbol{S}(\boldsymbol{\beta}_\tau)]$, that is, $\widehat{\boldsymbol{\beta}}_\tau \xrightarrow{p} \boldsymbol{\beta}_\tau$.

Next step is to derive the asymptotic normality of the GEEE estimator, $\widehat{\boldsymbol{\beta}}_\tau$. Consider $\boldsymbol{S}(\boldsymbol{\beta}_\tau) = \sum_{i=1}^{n} \boldsymbol{S}_i(\boldsymbol{\beta}_\tau)$, where $\boldsymbol{S}_i$ is a $pq \times 1$ vector defined as

$$\boldsymbol{S}_i = \begin{bmatrix} w_1 \sum_{t=1}^{m_i} \sum_{s=1}^{m_i} x_{it}^1 \sigma_{its\tau_1} \psi_{\tau_1}(\varepsilon_{it\tau_1}) \varepsilon_{it\tau_1} \\ \vdots \\ w_1 \sum_{t=1}^{m_i} \sum_{s=1}^{m_i} x_{it}^p \sigma_{its\tau_1} \psi_{\tau_1}(\varepsilon_{it\tau_1}) \varepsilon_{it\tau_1} \\ \vdots \\ w_q \sum_{t=1}^{m_i} \sum_{s=1}^{m_i} x_{it}^1 \sigma_{its\tau_q} \psi_{\tau_q}(\varepsilon_{it\tau_q}) \varepsilon_{it\tau_q} \\ \vdots \\ w_q \sum_{t=1}^{m_i} \sum_{s=1}^{m_i} x_{it}^p \sigma_{its\tau_q} \psi_{\tau_q}(\varepsilon_{it\tau_q}) \varepsilon_{it\tau_q} \end{bmatrix}$$

and let $Z_{ni} = \boldsymbol{\lambda}^{\mathsf{T}}(\boldsymbol{W} \otimes \boldsymbol{X}_i)^{\mathsf{T}} \boldsymbol{V}_{i\boldsymbol{\tau}}^{-1} \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}) \boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}$, where $\boldsymbol{\lambda}$ is a $pq \times 1$ unit vector, $\boldsymbol{\lambda}^{\mathsf{T}}\boldsymbol{\lambda} = 1$. The summands Z_{ni} are independent with $\mathbb{E}[Z_{ni}] = 0$ and $\text{Var}\left[n^{-1/2}\sum_{i=1}^{n} Z_{ni}\right] > \nu\prime > 0$, by condition **B2**. By the Minkowski's inequality, we have

$$
\begin{aligned}
\mathbb{E}|Z_{ni}|^{2+\nu} &= \mathbb{E}\left| \sum_{k=1}^{q}\sum_{l=1}^{p} w_k \lambda_{kl} \sum_{t=1}^{m_i}\sum_{s=1}^{m_i} x_{it}^{l} \sigma_{its\tau_k} \psi_{\tau_k}(\varepsilon_{it\tau_k}) \varepsilon_{it\tau_k} \right|^{2+\nu} \\
&= \mathbb{E}\left| \sum_{k=1}^{q}\sum_{l=1}^{p}\sum_{t=1}^{m_i}\sum_{s=1}^{m_i} w_k \lambda_{kl} x_{it}^{l} \sigma_{its\tau_k} \psi_{\tau_k}(\varepsilon_{it\tau_k}) \varepsilon_{it\tau_k} \right|^{2+\nu} \\
&\leq \left[\sum_{k=1}^{q}\sum_{l=1}^{p}\sum_{t=1}^{m_i}\sum_{s=1}^{m_i} \left(\mathbb{E}\left| w_k \lambda_{kl} x_{it}^{l} \sigma_{its\tau_k} \psi_{\tau_k}(\varepsilon_{it\tau_k}) \varepsilon_{it\tau_k} \right|^{2+\nu} \right)^{\frac{1}{2+\nu}} \right]^{2+\nu} \\
&\leq \left[\sum_{k=1}^{q}\sum_{l=1}^{p}\sum_{t=1}^{m_i}\sum_{s=1}^{m_i} |M \lambda_{kl} \sigma_{its\tau_k} \Delta^{\frac{1}{2+\nu}}| \right]^{2+\nu} \\
&\leq M^{2+\nu}\Delta \left[\sum_{k=1}^{q}\sum_{l=1}^{p} |\lambda_{kl}| \sum_{t=1}^{m_i}\sum_{s=1}^{m_i} |\sigma_{its\tau_k}| \right]^{2+\nu} \\
&\leq M^{2+\nu}\Delta \left[\sum_{k=1}^{q}\sum_{l=1}^{p} |\lambda_{kl}| \left\| \boldsymbol{V}_{i\tau_k}^{-1} \right\|_1 \right]^{2+\nu} \\
&\leq (Mm)^{2+\nu}\Delta |\boldsymbol{\lambda}^{\mathsf{T}} \mathbb{1}_{pq}|^{2+\nu} \\
&\leq (Mmpq)^{2+\nu}\Delta.
\end{aligned}
$$

Then, by the Liapounov CLT, **Theorem A.3**, $N^{-1/2}\boldsymbol{S}(\boldsymbol{\beta}_{\boldsymbol{\tau}})$ is a zero mean Gaussian vector with covariance matrix $\boldsymbol{D}_0(\boldsymbol{\tau})$.

Now consider the Taylor expansion of $N^{-1}\boldsymbol{S}$ in the neighborhood of $\boldsymbol{\beta}_{\boldsymbol{\tau}}$

$$
\begin{aligned}
N^{-1}\boldsymbol{S}(\boldsymbol{\beta}) \cong\ & N^{-1}\boldsymbol{S}(\boldsymbol{\beta}_{\boldsymbol{\tau}}) + \boldsymbol{D}_1(\boldsymbol{\beta}_{\boldsymbol{\tau}})(\boldsymbol{\beta} - \boldsymbol{\beta}_{\boldsymbol{\tau}}) \\
& - \left[N^{-1} \sum_{i=1}^{n} (\boldsymbol{W} \otimes \boldsymbol{X}_i)^{\mathsf{T}} \boldsymbol{V}_{i\boldsymbol{\tau}}^{-1} \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{\boldsymbol{\tau}}) (\mathbb{I}_q \otimes \boldsymbol{X}_i) + \boldsymbol{D}_1(\boldsymbol{\beta}_{\boldsymbol{\tau}}) \right] (\boldsymbol{\beta} - \boldsymbol{\beta}_{\boldsymbol{\tau}}).
\end{aligned}
$$

The last term of the equation is $o_p(1)$ by application of the LLN and the Slutsky's

Theorem. Because $\boldsymbol{S}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}}) = \mathbf{0}$ and $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}}$ is in the neighborhood of $\boldsymbol{\beta}_{\boldsymbol{\tau}}$, we have

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} - \boldsymbol{\beta}_{\boldsymbol{\tau}}) = \boldsymbol{D}_1^{-1}(\boldsymbol{\tau})\frac{1}{\sqrt{N}}\boldsymbol{S}(\boldsymbol{\beta}_{\boldsymbol{\tau}}) + o_p(1).$$

Therefore $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\Big(\mathbf{0}, \boldsymbol{D}_1^{-1}(\boldsymbol{\tau})\boldsymbol{D}_0(\boldsymbol{\tau})\boldsymbol{D}_1^{-1}(\boldsymbol{\tau})\Big)$. □

Proof of Theorem 2.3.4.

As for **Theorem 2.3.2**, the proof of **Theorem 2.3.4** consists in showing the convergence of $\widehat{\boldsymbol{D}}_1(\boldsymbol{\tau})$ and $\widehat{\boldsymbol{D}}_0(\boldsymbol{\tau})$. This is done following the same steps as **Theorem 2.3.2**. □

CHAPITRE III

DEUXIÈME ARTICLE : WEIGHTED ASYMMETRIC LEAST SQUARES REGRESSION FOR LONGITUDINAL DATA WITH FIXED-EFFECTS

Amadou Barry, Karim Oualkacha, Arthur Charpentier

Abstract.

The fixed-effects (FE) model is a commonly used model for the analysis of longitudinal data. FE is distinguished from other models, in part, by its ability to account for unmeasured factors. To improve and complete this approach, we adapted the FE model to expectile regression (ER). The ER captures the effect of covariates on the percentiles of the response variable. In this way, besides inheriting the attractive properties of the FE model, the expectile regression with fixed-effects (ERFE) makes it possible to study the influence of covariates on the location, scale and shape of the conditional distribution of the response. In our paper we study the asymptotic properties of the ERFE estimators and suggest robust estimators of their variances-covariances matrix. The simulation results show that the ERFE model estimators perform well and are competitive with the quantile regression with fixed-effects model estimators. Both methods are illustrated on a real data set.

Keywords : Expectile regression, Quantile regression, Fixed-effects, Within-transformation, Longitudinal data, Panel data.

3.1 Introduction

Ordinary least squares (OLS) regression is the most widely-used modelling method available in the toolkit of the data analyst. OLS and its extensions are used in all areas of applied science such as ecology, genetics, economics, sociology, medicine, to name a few. OLS regression models the average relationship between the covariates and the response variable. It summarizes their relationship by estimating the impact of the covariates on the conditional mean of the response variable. Unfortunately, this summary fails to account for the heterogeneity of the effects of the covariates on the response variable. For example, if the interest lies on the tails of the distribution of the response variable, OLS regression will not suffice, unless the effect of the covariates is the same on the location and the tail of the conditional distribution.

Today, with technological advances, there are alternative statistical modelling techniques developed to tackle this issue. One of them is the weighted asymmetric least squares regression, also called expectile regression (ER). ER pertains to the estimation of unknown expectiles of the cumulative distribution function of a response variable as a function of a set of covariates and a vector of regression coefficients. The vector coefficients represents the impact of the covariates on the location, scale, and shape of the response distribution. The ER method makes no assumption regarding the shape of the distribution of the response variable, allowing for investigation of a comprehensive class of covariate effects. In summary, ER provides a thorough and detailed insight into the influence of the covariates on the distribution of the response variable.

Expectile statistic is similar and related to quantile. Both characterize the distribution function from a sequence of asymmetric points. Expectile generalizes the mean and is computationally less demanding while quantile generalizes the median and is more robust. Another important point is that quantile is an order statistic while expectile is a weighted average. In other words, quantile focuses on the ordering of the observations while expectile targets their values (Farooq et Steinwart, 2017). In this paper, we are interested in the ER model, which is a natural generalization of the OLS regression, ER of level $\tau = 0.5$.

ER was first presented by Newey et Powell (1987). There is currently a growing interest in the literature regarding ER. The first study that revives the subject applied ER to spline and smoothing models (Schnabel et Eilers, 2009; Rossi et Harvey, 2009; Sobotka et Kneib, 2012; Sobotka *et al.*, 2013). Other research focused on contrasting ER and quantile regression (QR), showing how to derive quantiles from a fine grid of expectiles, on the crossing curves problem and on promoting application of ER (Kneib, 2013a; Schnabel et Eilers, 2013; Waltrup *et al.*, 2015a). Recently Schulze et Kauermann (2015) combined smoothing and random intercept models to fit clustered data with penalized splines.

Today ER is extended to many classes of models, Bayesian (Majumdar et Paul, 2016; Waldmann *et al.*, 2016; Xing et Qian, 2017), nonparametric (Righi *et al.*, 2014; Yang et Zou, 2015), nonlinear (Kim et Lee, 2016), neural network (Xu *et al.*, 2016; Jiang *et al.*, 2017) and support vector machine (Farooq et Steinwart, 2017). This paper contributes in this direction by extending ER to longitudinal/panel data.

Longitudinal study is by far the most valued observational study in applied research.

Longitudinal data arises in many applications such as epidemiology (Smith *et al.*, 2015), genetics (Furlotte *et al.*, 2012), econometrics (Hsiao, 2007), etc. Longitudinal studies consist of recording repeated measurements of characteristics of subjects of a cohort over a period of time. They capture the dynamic, development or change, at a group and subject level of the cohort. It allows the researcher to take into account the variability of the characteristic's measurements through time and the heterogeneity between subjects. With longitudinal studies, we obtain improved inference of parameters and more accurate predictions for individual outcomes. For further details about the advantages and limits of longitudinal data, see (Diggle *et al.*, 2013; Fitzmaurice *et al.*, 2008; Frees, 2004).

Because of the repeated measurements, longitudinal data is inherently correlated and its analysis thus requires appropriate statistical methods. Marginal and conditional models are the two main approaches used to analyze within-subject dependent data. Marginal models are designed to describe marginal means and the covariance structure of the data is treated as nuisance parameters. Therefore marginal models are specified like cross-sectional ones. The asymptotic properties of marginal ER estimators are presented by Barry *et al.* (2018a).

In conditional models the correlation structure is specified by the introduction of a subject-specific parameter in the model equation. The subject-specific parameter accounts for the within-subject dependency. It is common to all observations coming from the same subject and it represents the unobserved heterogeneity. More details about these family models can be found here (Stefanie *et al.*, 2016; Lee et Nelder, 2004).

In econometrics, there are two popular conditional models : the fixed-effects and the random effects models (Greene, 2012). The fixed-effects (FE) model is a commonly used model to analyze longitudinal data, mainly because it considers the correlation between the covariates of the model and the unmeasured factors. For example, the ability of the subject which is correlated to the education level is always a missing covariate in the Mincer earnings equation model (Lemieux, 2006). In health studies, the birthweight effect on infant mortality depends on the genetic characteristics of the mother ; usually unmeasured.

In this paper, we combine the asymmetric least squares regression to the FE model to allow the FE model to estimate the effects of the covariates on the location, scale and shape of the distribution of the response variable. Our main contribution is the derivation of the asymptotic properties of the ER with fixed-effects (ERFE) estimators and the adaptation of the within-transformation strategy in the ER framework to solve the incidental parameter problem. Take note that the quantile regression method was applied in the FE framework (Koenker, 2004; Galvao, 2011). In that context, these researchers were unable to apply the within-transformation strategy to solve the incidental parameter problem. Instead, they rely on the Bahadur representation of the infinite subject-specific parameter.

In the next Section 3.2, we introduce the expectile function and the expectile regression model and then present the expectile regression with fixed-effects (ERFE) model. In Section 3.3, we derive the asymptotic properties of the ERFE estimator and propose an estimator of its variance-covariance (VC) matrix. We present the small sample performance of the ERFE estimator in Section 3.4 and its application to a real dataset in Section 3.5. The conclusion is in Section 3.6 and the proofs are

in the Appendix.

3.2 Models and Methods

3.2.1 Expectile and expectile regression

Expectile of level $\tau \in [0,1]$ of a r.v Y is defined as the unique solution of

$$\mu_\tau(Y) = \underset{\theta \in \mathbb{R}}{\operatorname{argmin}}\, \mathbb{E}\{\rho_\tau(Y-\theta)\},$$

where $\rho_\tau(t) = |\tau - \mathbb{1}(t \leq 0)| \cdot t^2$ is the asymmetric square loss function that assigns weights τ and $1-\tau$ to positive and negative deviations respectively.
Expectile function has attractive properties. Expectile generalizes the mean, that is the mean of a r.v Y is equal to $\mu_{0.5}(Y)$. Expectile is location and scale equivariant, that is for $s > 0$ and $t \in \mathbb{R}$, $\mu_\tau(sY+t) = s\mu_\tau(Y) + t$. Most importantly, expectile function summarizes the cumulative distribution function (c.d.f.) of a r.v in the same way that quantile function does. A detailed study of the other properties of the expectile function can be found in Newey et Powell (1987).

The main limit raised about expectile function is its interpretation. But, this should not be considered a serious disadvantage. We have found that expectile function is a weighted average :

$$\mu_\tau(Y) = \mu_\tau = \mathbb{E}\left[\frac{\psi_\tau(Y-\mu_\tau)}{\mathbb{E}\left[\psi_\tau(Y-\mu_\tau)\right]}Y\right],$$

where $\psi_\tau(t) = |\tau - \mathbb{1}(t \leq 0)|$ is the check function. The only subtlety is that the weights are random. Given a random sample, $\{(y_i)\}_{i=1}^n$, the corresponding τ-

th sample expectile

$$\widehat{\mu}_\tau = \sum_{i=1}^{n} \frac{\psi_\tau(y_i - \widehat{\mu}_\tau)}{\sum_{l=1}^{n} \psi_\tau(y_l - \widehat{\mu}_\tau)} y_i \tag{3.1}$$

is the weighted mean, where the weights depend on the sample data. For a fixed θ, equation (3.1) is derived as the solution which minimizes the empirical loss function

$$\frac{1}{n} \sum_{i=1}^{n} \rho_\tau(y_i - \theta). \tag{3.2}$$

Newey et Powell (1987) introduced the expectile function to the linear regression as a tool to study the effects of covariates on the location, scale and shape of the distribution of the response variable. Consider the linear regression model

$$y_i = \boldsymbol{x}_i{}^\mathsf{T} \boldsymbol{\beta} + \varepsilon_i \quad \text{with} \quad \mu_\tau(\varepsilon_i) = 0, \tag{3.3}$$

where $\boldsymbol{x}_i$ is a $p \times 1$ vector of covariates, y_i and ε_i are respectively the response variable and the random error with unspecified distribution function. $\boldsymbol{\beta} \in \mathbb{R}^p$ is the unknown parameter that needs to be estimated. The assumption $\mu_\tau(\varepsilon_i) = 0$ ensures that the random error is centred on the τ-th expectile. In the linear regression settings, the corresponding linear ER model for a fixed $\tau \in (0, 1)$ is given as

$$\mu_\tau(y_i | \boldsymbol{x}_i) = \boldsymbol{x}_i{}^\mathsf{T} \boldsymbol{\beta}_\tau. \tag{3.4}$$

The ER estimator is defined as the vector $\widehat{\boldsymbol{\beta}}_\tau$ which minimizes the function

$$\sum_{i=1}^{n} \rho_\tau\left(y_i - \boldsymbol{x}_i{}^\mathsf{T} \boldsymbol{\beta}_\tau\right)$$

over $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$. Since the loss function $\rho_\tau(t)$ is continuously differentiable, we have through the first order condition

$$\widehat{\boldsymbol{\beta}}_\tau = \left[\sum_{i=1}^{n} \psi_\tau(y_i - \boldsymbol{x}_i{}^\mathsf{T} \widehat{\boldsymbol{\beta}}_\tau) \boldsymbol{x}_i \boldsymbol{x}_i{}^\mathsf{T} \right]^{-1} \sum_{i=1}^{n} \psi_\tau(y_i - \boldsymbol{x}_i{}^\mathsf{T} \widehat{\boldsymbol{\beta}}_\tau) \boldsymbol{x}_i y_i \tag{3.5}$$

where $\psi_\tau(t) = |\tau - \mathbb{1}(t \leq 0)|$ is the check function. The ER estimator can be computed as iterated weighted least squares estimators. For the special case of $\tau = 0.5$, $\widehat{\boldsymbol{\beta}}_{0.5}$ is the OLS regression estimator. This makes ER a natural complement of the OLS regression. Whenever the conditional mean is the target, the ER should be used in order to have an overall picture of the impact of the covariate.

In addition, expectile is an interesting alternative to quantile. It is shown that expectile of level τ is a quantile of level α, the expression of α can be found here (Waltrup *et al.*, 2015a). Empirically one can obtain a rough idea concerning the location of $\widehat{\boldsymbol{\beta}}_\tau$ in the conditional distribution of y_i given $\boldsymbol{x}_i$ by calculating the proportion of observations for which $y_i \leq \boldsymbol{x}_i^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_\tau$. As noted by Newey et Powell (1987), in the case where $\boldsymbol{x}_i$ is independent of ε_i this proportion will be a consistent estimator of $F(\mu_\tau(\varepsilon_i|\boldsymbol{x}_i))$, where $F(\cdot)$ is the c.d.f. of ε_i and $\mu_\tau(\varepsilon_i|\boldsymbol{x}_i)$ is the τ-th expectile of ε_i. Expectile is more advantageous computationally than quantile. Another important advantage of the ER estimator is its variance-covariance matrix is readily available to construct confidence intervals and hypothesis tests, whereas that of the QR estimator requires estimating the unknown density at 0, a computationally difficult and expensive task that can cause some numerical computational problems, (Yin et Cai, 2005).

The ER estimator can be derived through a likelihood-based approach by assuming an asymmetric normal distribution (AND) for the disturbance (Aigner *et al.*, 1976). Assume $\varepsilon_i \sim \text{AND}(\varepsilon; \mu, \sigma^2, \tau)$ with density

$$f(\varepsilon; \mu, \sigma^2, \tau) = \frac{2}{\sqrt{\pi\sigma^2}} \frac{\sqrt{\tau(1-\tau)}}{\sqrt{\tau} + \sqrt{1-\tau}} \exp\left\{ -\rho_\tau\left(\frac{\varepsilon - \mu}{\sigma}\right)\right\},$$

where σ is a known nuisance parameter. Let $\mu_{i\tau} = \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta}_\tau$ and assume the n observations are independent then the ER estimator is equivalent to the maximum of the

likelihood

$$\mathbb{L}(\boldsymbol{\beta}_\tau; \sigma, \tau, \boldsymbol{y}) \propto \sigma^{-2n} \exp\left\{ -\sum_{i=1}^{n} \rho_\tau \left(\frac{y_i - \boldsymbol{x}_i^\mathsf{T} \boldsymbol{\beta}_\tau}{\sigma} \right) \right\},$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^\mathsf{T}$.

3.2.2 Fixed-effects model for longitudinal data

Consider the standard linear fixed-effects model for longitudinal data

$$y_{ij} = \boldsymbol{x}_{ij}^\mathsf{T} \boldsymbol{\beta} + \alpha_i + \varepsilon_{ij}, \tag{3.6}$$

where y_{ij} is the scalar response variable, $\boldsymbol{x}_{ij} = (x_{ij}^1, x_{ij}^2, \ldots, x_{ij}^p)^\mathsf{T} \in \mathbb{R}^p$ is the vector of covariates measured on subject i at time j, α_i is the subject-specific effects parameter and ε_{ij} is a random error with unspecified distribution function. Equation model (3.6) is conveniently represented in individual notation as

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_i, \tag{3.7}$$

where $\boldsymbol{y}_i$ and $\boldsymbol{\varepsilon}_i$ are $m \times 1$ vectors, $\boldsymbol{X}_i$ is a $m \times p$ design matrix and $\boldsymbol{Z}_i$ is a $m \times n$ incidence matrix and $\boldsymbol{\alpha}$ is a $n \times 1$ subject-specific effects vector. Equation (3.7) can also be represented in matrix form by stacking all the observations

$$\boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{Z} \boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{3.8}$$

where $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ are $N \times 1$ vectors, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are respectively $N \times p$ and $N \times n$ matrix with $N = mn$. The incidence matrix $\boldsymbol{Z}$ identifies the n distinct subjects of the sample. Take note that a balanced design is implicitly assumed.

Given the complications introduced by the presence of the infinite dimensional parameter $\boldsymbol{\alpha}$ in the fixed-effects model, various strategies have been proposed to estimate

the parameters of the model. The most frequently applied strategy is the within-transformation method. It consists of pre-multiplying both sides of the equation model (3.8) by a $N \times N$ idempotent matrix $\boldsymbol{M_Z} = \mathbb{I}_N - \boldsymbol{Z}(\boldsymbol{Z}^\mathsf{T}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\mathsf{T}$ to get rid of the infinite-dimensional parameter. The model that results from this transformation is represented as

$$\boldsymbol{y}^* = \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*, \tag{3.9}$$

where $\boldsymbol{y}^* = \boldsymbol{M_Z}\boldsymbol{y}$ and $\boldsymbol{X}^*$ and $\boldsymbol{\varepsilon}^*$ are defined similarly. Subsequently, we fit the ordinary least squares (OLS) regression to the transformed data and derive the estimator of the parameter of interest. The OLS estimator of the fixed-effects model, known as the within-transformation estimator, is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{y}^*. \tag{3.10}$$

The within-transformation estimator is consistent and asymptotically normally distributed, (Greene, 2012). For further details about the fixed-effects model and the within-transformation strategy, see (Rabe-Hesketh et Skrondal, 2008; Wooldridge, 2002). In the following subsection, we study the expectile regression for fixed-effects (ERFE) model and derive the iterative-within-transformation ERFE estimator.

3.2.3 ERFE model for longitudinal data

We define the ERFE model of the linear fixed-effects model (3.6) as

$$\mu_\tau(y_{ij}|\alpha_i, \boldsymbol{x}_{ij}) = \boldsymbol{x}_{ij}{}^\mathsf{T}\boldsymbol{\beta}_\tau + \boldsymbol{z}_{ij}{}^\mathsf{T}\boldsymbol{\alpha}, \tag{3.11}$$

for every fixed $\tau \in (0,1)$. It is implicitly assumed that $\mu_\tau(\varepsilon_{ij}|\alpha_i, \boldsymbol{x}_{ij}) = 0$. This assumption ensures that the random error is centred on the τ-th expectile. In this setting the parameter $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$ captures the influence of the covariates $\boldsymbol{x}_{ij}$ on the

location, scale and shape of the conditional distribution of the response variable y_{ij}. The subject-specific effects parameter $\boldsymbol{\alpha}$ is assumed to be independent of τ across the percentiles and to have a pure location-shift effect on the conditional expectile of the response. Assuming a τ-dependency of the subject-specific effects implies estimating its distribution with the number m of within-subject observations, which is relatively small in most applications. Take note also that no assumption is made as to the shape of the distribution of the response variable.

The ERFE estimator of the parameter of the model (3.11) is defined as the vector minimizing the following objective function

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\rho_\tau\big(y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta}_\tau - \boldsymbol{z}_{ij}{}^{\mathsf{T}}\boldsymbol{\alpha}\big). \tag{3.12}$$

Since the loss function $\rho_\tau(\cdot)$ is continuously differentiable, we can apply the first-order condition. The resulting ERFE estimator $\widehat{\boldsymbol{\beta}}_\tau$ can be defined as

$$\widehat{\boldsymbol{\beta}}_\tau = \Big\{\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_\tau\big[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{Z}\widehat{\boldsymbol{\alpha}}\big]\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}\Big\}^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_\tau\big[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{Z}\widehat{\boldsymbol{\alpha}}\big]\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)\boldsymbol{y} \tag{3.13}$$

where the diagonal check function matrix is

$$\boldsymbol{\Psi}_\tau\big[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{Z}\widehat{\boldsymbol{\alpha}}\big] = \operatorname{diag}\Big(\psi_\tau(y_{11}-\boldsymbol{x}_{11}{}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{z}_{11}{}^{\mathsf{T}}\widehat{\boldsymbol{\alpha}}),\ldots,\psi_\tau(y_{nm}-\boldsymbol{x}_{nm}{}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{z}_{nm}{}^{\mathsf{T}}\widehat{\boldsymbol{\alpha}})\Big). \tag{3.14}$$

The projection matrix $\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)$ and its complement $\widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau)$ are idempotent matrices and are defined as

$$\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau) = \mathbb{I}_N - \widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau) \text{ and } \widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau) = \boldsymbol{Z}(\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{\Psi}_\tau\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{\Psi}_\tau.$$

The function $\boldsymbol{\Psi}_\tau(\cdot)$ defined in (3.14) depends on the subject-specific parameter estimator $\widehat{\boldsymbol{\alpha}}$ which, by the first-order condition of (3.12), verifies the relationship

$$\boldsymbol{Z}\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau)(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\tau). \tag{3.15}$$

Using equation (3.15), the argument of the check function matrix can be written as

$$\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{Z}\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\tau) \tag{3.16}$$

and then eliminates the incidental parameter estimate from expression (3.13) of the ERFE estimator.

Now, using the following relationship

$$\boldsymbol{\Psi}_\tau\big[\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\tau)\big]\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau) = \widehat{\boldsymbol{M}}_{\boldsymbol{Z}}{}^{\mathsf{T}}(\tau)\boldsymbol{\Psi}_\tau\big[\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\tau)\big],$$

we can redefine the ERFE estimator as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_\tau = &\Big\{\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_\tau\big[\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\tau)\big]\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}\Big\}^{-1} \\ &\times \widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_\tau\big[\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\tau)\big]\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)\boldsymbol{y}.\end{aligned} \tag{3.17}$$

As shown above, the ERFE estimator is a combination of the ER estimator (3.5) and the within-transformation estimator (3.10). An ER estimator : because the expression of the ERFE estimator depends on the check function matrix which is updated at each iteration until convergence. A within-transformation estimator : because we can eliminate the incidental parameter from the expression of the ERFE estimator by pre-multiplying the initial data $[\boldsymbol{y}, \boldsymbol{X}]$ by the projection matrix $\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)$ and then applying the ER method on the transformed data.

In summary, the within-transformation strategy is extended to the ER framework. The strategy is derived by applying the projection matrix $\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)$ to the initial data $[\boldsymbol{y}, \boldsymbol{X}]$, in order to eliminate the subject-specific effects parameter. In addition, like the ER estimator (3.5), the within ERFE estimator is computed iteratively, using the iterative weighted least squares algorithm. The detailed algorithm for computing the iterative-within-transformation ERFE estimator is summarized in the following stepwise procedures.

Algorithm : The iterative within-transformation ERFE algorithm

---

Step 1. For fixed τ, let $\widetilde{\boldsymbol{\beta}}_{\tau}^{(0)} = \widehat{\boldsymbol{\beta}}_{\tau}$, the ER estimator of Newey et Powell (1987) and $\widehat{\varepsilon}_{ij\tau}^{(0)} = \widetilde{y}_{ij}^{(0)} - \widetilde{\boldsymbol{x}}_{ij}^{(0)\mathsf{T}}\widetilde{\boldsymbol{\beta}}_{\tau}^{(0)}$ for $i \in \{1, \ldots, n\}$, and $j \in \{1, \ldots, m\}$.

Step 2. Given $\widetilde{\boldsymbol{\beta}}_{\tau}^{(r-1)}$ from the $r-1$ step, update

$$\begin{aligned}
&\widehat{\boldsymbol{P}}_{\boldsymbol{Z}}^{(r)}(\tau) \leftarrow \boldsymbol{Z}(\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}}_{\tau}^{*(r-1)})\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}}_{\tau}^{*(r-1)}),\\
&\widetilde{\boldsymbol{y}^{*}}^{(r)} \leftarrow \widehat{\boldsymbol{M}}_{\boldsymbol{Z}}^{(r)}(\tau)\boldsymbol{y},\\
&\widetilde{\boldsymbol{X}^{*}}^{(r)} \leftarrow \widehat{\boldsymbol{M}}_{\boldsymbol{Z}}^{(r)}(\tau)\boldsymbol{X}
\end{aligned}$$

where $\widehat{\boldsymbol{\varepsilon}}_{\tau}^{*(r-1)} = \widetilde{\boldsymbol{y}^{*}}^{(r-1)} - \widetilde{\boldsymbol{X}^{*}}^{(r-1)}\widetilde{\boldsymbol{\beta}}_{\tau}^{(r-1)}$.

Step 3. Update $\widehat{\boldsymbol{\beta}}_{\tau}^{(r)}$ by

$$\widehat{\boldsymbol{\beta}}_{\tau}^{(r)} \leftarrow \widehat{\boldsymbol{\beta}}_{\tau}^{(r-1)} + \left[\widetilde{\boldsymbol{X}^{*}}^{(r-1)\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}}_{\tau}^{*(r-1)})\widetilde{\boldsymbol{X}^{*}}^{(r-1)}\right]^{-1}\widetilde{\boldsymbol{X}^{*}}^{(r-1)\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}}_{\tau}^{*(r-1)})\widehat{\boldsymbol{\varepsilon}}_{\tau}^{*(r-1)}.$$

Step 4. Update $\widehat{\boldsymbol{\varepsilon}}_{\tau}^{*(r)}$ by

$$\widehat{\boldsymbol{\varepsilon}}_{\tau}^{*(r)} \leftarrow \widetilde{\boldsymbol{y}^{*}}^{(r)} - \widetilde{\boldsymbol{X}^{*}}^{(r)}\widehat{\boldsymbol{\beta}}_{\tau}.$$

Step 5. Repeat the above iteration, Step 2-4, until convergence.

When $\tau = 0.5$ we have $\boldsymbol{\Psi}_{\tau} = 0.5\mathbb{I}_N$ and the iterative within-transformation ERFE estimator is nothing else than the OLS within-transformation estimator (3.9).

From the above algorithm, the multiplication of a vector (say $\boldsymbol{y}$) by the matrix $\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)$ deviates that vector from its expectile as shown by the following expression

$$\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)\boldsymbol{y} = \Bigg(y_{11} - \sum_{j=1}^{m} \frac{\psi_\tau(\widehat{\varepsilon}_{1j})}{\sum_{k=1}^{m} \psi_\tau(\widehat{\varepsilon}_{1k})} y_{1j}, \ldots, y_{1m} - \sum_{j=1}^{m} \frac{\psi_\tau(\widehat{\varepsilon}_{1j})}{\sum_{k=1}^{m} \psi_\tau(\widehat{\varepsilon}_{1k})} y_{1j}, \ldots,$$
$$y_{n1} - \sum_{j=1}^{m} \frac{\psi_\tau(\widehat{\varepsilon}_{nj})}{\sum_{k=1}^{m} \psi_\tau(\widehat{\varepsilon}_{nk})} y_{nj}, \ldots, y_{nm} - \sum_{j=1}^{m} \frac{\psi_\tau(\widehat{\varepsilon}_{nj})}{\sum_{k=1}^{m} \psi_\tau(\widehat{\varepsilon}_{nk})} y_{nj}\Bigg)^{\mathsf{T}}.$$

We can see, from this expression, how the projection matrix $\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)$ eliminates the subject-specific effects parameter and any other time-invariant covariates from the initial model. For a matrix, like the design matrix $\boldsymbol{X}$, the transformation is applied column-wise.

ERFE model for a sequence of expectiles

The preceding development shows that the within-transformation strategy applied to concentrate out the incidental parameter in the classical linear fixed-effects model can be generalized in the ERFE framework. Now, we present the ERFE estimator for a sequence of expectiles using the transformed data. The sequence of expectiles, for example the mean and a few expectiles above and below it, is necessary in the description of the conditional distribution of the response variable. In addition, the simultaneous estimation allows the multiple expectiles to share strength among each other and to gain better estimation accuracy than individually estimated expectile functions (Liu et Wu, 2011).

The iterative within-transformation ERFE estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} = [\widehat{\boldsymbol{\beta}}_{\tau_1}{}^{\mathsf{T}}, \ldots, \widehat{\boldsymbol{\beta}}_{\tau_q}{}^{\mathsf{T}}]^{\mathsf{T}}$, for a sequence of percentiles $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)^{\mathsf{T}}$, is defined as the minimum of the following

objective function

$$\sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \rho_{\tau_k}\Big(y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta}_{\tau_k} - \boldsymbol{z}_{ij}{}^{\mathsf{T}}\boldsymbol{\alpha}\Big). \tag{3.18}$$

The vector $\boldsymbol{v} = (v_1, \ldots, v_q)^{\mathsf{T}}$ is the vector of weights controlling the relative influence of the q expectiles $\{\tau_1, \ldots, \tau_q\}$. For a sequence of expectiles, the iterative within-transformation ERFE estimator is defined as

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} = \Big[(\mathbb{I}_q \otimes \widehat{\boldsymbol{X}^*})^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}})(\boldsymbol{V} \otimes \widehat{\boldsymbol{X}^*})\Big]^{-1}(\boldsymbol{V} \otimes \widehat{\boldsymbol{X}^*})^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}})(\mathbb{1}_q \otimes \widehat{\boldsymbol{y}^*}) \tag{3.19}$$

where $\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}}) = \Big[\operatorname{diag}\big(\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{y}^*} - \widehat{\boldsymbol{X}^*}\widehat{\boldsymbol{\beta}}_{\tau_k})\big)\Big]_{k=1}^{q}$, $\boldsymbol{V} = [\operatorname{diag}(v_k)]_{k=1}^{q}$ and the transformed data $[\widehat{\boldsymbol{y}^*}, \widehat{\boldsymbol{X}^*}]$ is obtained by pre-multiplying the matrix $\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\boldsymbol{\tau})$ to the initial data $[\boldsymbol{y}, \boldsymbol{X}]$. The projection matrix is defined as

$$\widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\boldsymbol{\tau}) = (\boldsymbol{v} \otimes \boldsymbol{Z})\Big[(\boldsymbol{v} \otimes \boldsymbol{Z})^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}})(\mathbb{1}_q \otimes \boldsymbol{Z})\Big]^{-1}(\mathbb{1}_q \otimes \boldsymbol{Z})^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}}),$$

and $\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\boldsymbol{\tau}) = \mathbb{I}_{nmq} - \widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\boldsymbol{\tau})$.

However, the within-transformation method is not applicable in the QR framework because quantile does not commute with linear transformation like expectations. Instead, Koenker (2004) relies on Bahadur's representation to concentrate out the nuisance parameter.

3.3 Asymptotics

In this section, the asymptotic properties of the ERFE estimator are presented. As stated by Koenker (2004), the presence of the incidence parameter, whose dimension tends to infinity, can raise other problems. For this reason, we present first the asymptotic results of the ERFE estimator in the simplest case, namely for a single τ.

We then present the asymptotic properties of the ERFE estimator for a simultaneous sequence of percentiles $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)$. The section ends with the suggestion of an estimator of the VC-matrix of the ERFE estimator. All the proofs are available in the Appendix.

Asymptotics for a single expectile

In the following the asymmetric square-loss function of the ERFE model,

$$\rho_\tau\Big(y_{ij} - \boldsymbol{x}_{ij}{}^\top\boldsymbol{\beta}_\tau - \boldsymbol{z}_{ij}{}^\top\boldsymbol{\alpha}\Big),$$

is replaced by the equivalent, in term of optimization, loss function

$$\rho_\tau\Big(y_{ij} - \mu_{ij\tau} - \boldsymbol{x}_{ij}{}^\top\boldsymbol{\delta}_{1\tau}/\sqrt{nm} - \boldsymbol{z}_{ij}{}^\top\boldsymbol{\delta}_0/\sqrt{m}\Big) - \rho_\tau(y_{ij} - \mu_{ij\tau}),$$

where $\mu_{ij\tau} = \boldsymbol{x}_{ij}{}^\top\boldsymbol{\beta}_\tau + \boldsymbol{z}_{ij}{}^\top\boldsymbol{\alpha}$. Now, observe that the following estimator

$$\widehat{\boldsymbol{\delta}} = \begin{pmatrix} \widehat{\boldsymbol{\delta}}_0 \\ \widehat{\boldsymbol{\delta}}_{1\tau} \end{pmatrix} = \begin{pmatrix} \sqrt{m}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \\ \sqrt{nm}\Big[\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau\Big] \end{pmatrix}$$

minimizes the new objective function

$$R_{nm}(\boldsymbol{\delta}) = \sum_{i=1}^{n}\sum_{j=1}^{m}\rho_\tau\Big(y_{ij} - \mu_{ij\tau} - \boldsymbol{x}_{ij}{}^\top\boldsymbol{\delta}_{1\tau}/\sqrt{nm} - \boldsymbol{z}_{ij}{}^\top\boldsymbol{\delta}_0/\sqrt{m}\Big) - \rho_\tau(y_{ij} - \mu_{ij\tau}). \tag{3.20}$$

The asymptotic theory of the ERFE estimator is derived using this new objective function (3.20) and under the following assumptions

A1. The data $\{(\boldsymbol{y}_i, \boldsymbol{X}_i)\}_{i=1}^n$ are independent across i, and

$\mathrm{Var}\Big[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\Big] = \mathbb{E}\Big[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\boldsymbol{\varepsilon}_{i\tau}{}^\top\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\Big] = \boldsymbol{\Sigma}_{i\tau}$, where

$\boldsymbol{\varepsilon}_{i\tau} = (\varepsilon_{i1\tau}, \ldots, \varepsilon_{im\tau})^\top$, $\varepsilon_{ij\tau} = y_{ij} - \boldsymbol{x}_{ij}{}^\top\boldsymbol{\beta}_\tau$ and $\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}) = \Big[\mathrm{diag}(\psi_\tau(\varepsilon_{ij\tau}))\Big]_{j=1}^m$.

A2. The limiting forms of the following matrices are positive definite

$$\boldsymbol{D}_0(\tau) = \lim_{\substack{m\to\infty \\ n\to\infty}} m^{-1} \begin{pmatrix} \boldsymbol{Z}^\top \boldsymbol{\Sigma}_\tau \boldsymbol{Z} & \boldsymbol{Z}^\top \boldsymbol{\Sigma}_\tau \boldsymbol{X}/\sqrt{n} \\ \boldsymbol{X}^\top \boldsymbol{\Sigma}_\tau \boldsymbol{Z}/\sqrt{n} & \boldsymbol{X}^\top \Sigma_\tau \boldsymbol{X}/n \end{pmatrix},$$

$$\boldsymbol{D}_1(\tau) = \lim_{\substack{m\to\infty \\ n\to\infty}} m^{-1} \begin{pmatrix} \boldsymbol{Z}^\top \mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{Z} & \boldsymbol{Z}^\top \mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}/\sqrt{n} \\ \boldsymbol{X}^\top \mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{Z}/\sqrt{n} & \boldsymbol{X}^\top \mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}/n \end{pmatrix},$$

where $\boldsymbol{\Sigma}_\tau = \mathrm{Var}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau] = \mathrm{diag}[\boldsymbol{\Sigma}_{i\tau}]_{i=1}^n$.

A3. $\max_{i,j} \|\boldsymbol{x}_{ij}\| < M$, where M is a constant.

The stated assumptions **A1**-**A3** are standard for longitudinal data models. Condition **A1** ensures independence across individuals, but allows a within-subject dependency and heterogeneity across individuals. Condition **A2** is a full rank condition and is used to invoke the Lindeberg-Feller Central Limit Theorem. Observe that, when $\tau = 1/2$ then $\boldsymbol{D}_1(\tau)$ simplifies and **A2** reduces to a condition on the matrices $\boldsymbol{X}^\top\boldsymbol{X}/nm$ and $\boldsymbol{Z}^\top\boldsymbol{Z}/m$. Condition **A3** is useful both for the application of the Lindeberg-Feller Central Limit Theorem and for ensuring the finite dimensional convergence of the objective function.

Theorem 3.3.1. *Assume conditions **A1**-**A3**, with $n, m \to \infty$, and $\mathbb{E}|\psi_\tau(\varepsilon_{ij\tau})|^{4+\nu} < \Delta < \infty$ and $\mathbb{E}|\varepsilon_{ij\tau}|^{4+\nu} < \Delta < \infty$ for some $\nu > 0$. Then $\widehat{\boldsymbol{\delta}}_{1\tau}$ the components of the minimizer, $\widehat{\boldsymbol{\delta}}$, converge in distribution to a Gaussian random vector with mean zero and variance-covariance matrix given by the lower right $p \times p$ block of the matrix $\boldsymbol{D}_1^{-1}(\tau)\boldsymbol{D}_0(\tau)\boldsymbol{D}_1^{-1}(\tau)$. In others words*

$$\sqrt{nm}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau) \xrightarrow{d} \mathcal{N}\Big(0, \Big[\boldsymbol{D}_1^{-1}(\tau)\boldsymbol{D}_0(\tau)\boldsymbol{D}_1^{-1}(\tau)\Big]_{22}\Big).$$

To show the closed form of the above matrix $\left[\boldsymbol{D}_1^{-1}(\tau)\boldsymbol{D}_0(\tau)\boldsymbol{D}_1^{-1}(\tau)\right]_{22}$ assume that the limiting forms of the following matrices are positive definite

$$\widetilde{\boldsymbol{D}}_0(\tau) = \lim_{\substack{m\to\infty \\ n\to\infty}} \boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}(\tau)^{\mathsf{T}}\boldsymbol{\Sigma}_\tau\boldsymbol{M}_{\boldsymbol{Z}}(\tau)\boldsymbol{X},$$

$$\widetilde{\boldsymbol{D}}_1(\tau) = \lim_{\substack{m\to\infty \\ n\to\infty}} \boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}(\tau)\,\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon})]\boldsymbol{M}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}$$

where $\boldsymbol{M}_{\boldsymbol{Z}}(\tau) = \mathbb{I} - \boldsymbol{P}_{\boldsymbol{Z}}(\tau)$ and $\boldsymbol{P}_{\boldsymbol{Z}}(\tau) = \boldsymbol{Z}\left[\boldsymbol{Z}^{\mathsf{T}}\,\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon})]\boldsymbol{Z}\right]^{-1}\boldsymbol{Z}^{\mathsf{T}}\,\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon})]$.

Under the above conditions and the conditions of **Theorem 3.3.1** it follows that

Lemma 3.3.1.1.

$$\left[\boldsymbol{D}_1^{-1}(\tau)\boldsymbol{D}_0(\tau)\boldsymbol{D}_1^{-1}(\tau)\right]_{22} = \widetilde{\boldsymbol{D}}_1^{-1}(\tau)\widetilde{\boldsymbol{D}}_0(\tau)\widetilde{\boldsymbol{D}}_1^{-1}(\tau).$$

Asymptotics for several expectiles

In this section, the asymptotic properties of the ERFE estimator for several expectiles $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_q)$ are derived using the transformed data, $[\boldsymbol{y}^*; \boldsymbol{X}^*]$, where $\boldsymbol{X}^* = \boldsymbol{M}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}$ and $\boldsymbol{y}^* = \boldsymbol{M}_{\boldsymbol{Z}}(\tau)\boldsymbol{y}$. Both projection matrices $\boldsymbol{M}_{\boldsymbol{Z}}(\tau)$ and $\boldsymbol{P}_{\boldsymbol{Z}}(\tau)$ are idempotent and are defined as

$$\boldsymbol{M}_{\boldsymbol{Z}}(\tau) = \mathbb{I}_N - \boldsymbol{P}_{\boldsymbol{Z}}(\tau) \text{ and } \boldsymbol{P}_{\boldsymbol{Z}}(\tau) = \boldsymbol{Z}(\boldsymbol{Z}^{\mathsf{T}}\,\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}^*_\tau)]\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\mathsf{T}}\,\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}^*_\tau)]$$

where $\boldsymbol{\varepsilon}^*_\tau = \boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{\beta}_\tau$.

A robust estimator of the VC-matrix is also proposed. Assume the following conditions

B1. The data $\{(\boldsymbol{y}_i, \boldsymbol{X}_i)\}_{i=1}^n$ are independent across i and

$\text{Var}\left[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}^*_{i\boldsymbol{\tau}})\boldsymbol{\varepsilon}^*_{i\boldsymbol{\tau}}\right] = \mathbb{E}\left[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}^*_{i\boldsymbol{\tau}})\boldsymbol{\varepsilon}^*_{i\boldsymbol{\tau}}\boldsymbol{\varepsilon}^{*\mathsf{T}}_{i\boldsymbol{\tau}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}^*_{i\boldsymbol{\tau}})\right] = \boldsymbol{\Sigma}^*_{i\boldsymbol{\tau}}$, where $\boldsymbol{\varepsilon}^*_{i\boldsymbol{\tau}} = \left(\boldsymbol{\varepsilon}^{*\mathsf{T}}_{i\tau_1}, \ldots, \boldsymbol{\varepsilon}^{*\mathsf{T}}_{i\tau_q}\right)^{\mathsf{T}}$

$\boldsymbol{\varepsilon}^*_{i\tau_k} = (\varepsilon^*_{i1\tau_k}, \ldots, \varepsilon^*_{im\tau_k})^{\mathsf{T}}$, $\varepsilon^*_{ij\tau_k} = y^*_{ij} - \boldsymbol{x}^{*\mathsf{T}}_{ij}\boldsymbol{\beta}_{\tau_k}$ and $\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}^*_{i\boldsymbol{\tau}}) = \left[\text{diag}(\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{i\tau_k}))\right]_{k=1}^q$.

B2. The limiting forms of the following matrices are positive definite

$$\boldsymbol{D}_0(\boldsymbol{\tau}) = \lim_{n\to\infty} \; (\boldsymbol{V}\otimes\boldsymbol{X}^*)^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}})\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}\boldsymbol{\varepsilon}^{*\mathsf{T}}_{\boldsymbol{\tau}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}})](\boldsymbol{V}\otimes\boldsymbol{X}^*)/nm.$$

$$\boldsymbol{D}_1(\boldsymbol{\tau}) = \lim_{n\to\infty} \; (\mathbf{I}_q\otimes\boldsymbol{X}^*)^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}})](\boldsymbol{V}\otimes\boldsymbol{X}^*)/nm$$

B3. $\max_{\substack{1\le i\le n\\ 1\le j\le m}} \left\|x^*_{ij}\right\| < M$, where M is a constant.

Theorem 3.3.2. *Suppose conditions **B1-B3** are satisfied, and that $n, m \to \infty$. If $\mathbb{E}|\psi_\tau(\varepsilon^*_{ij\tau})|^{4+\nu} < \Delta < \infty$ and $\mathbb{E}|\varepsilon^*_{ij\tau}|^{4+\nu} < \Delta < \infty$ then*

$$\sqrt{nm}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} - \boldsymbol{\beta}_{\boldsymbol{\tau}}) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{D}_1^{-1}(\boldsymbol{\tau})\boldsymbol{D}_0(\boldsymbol{\tau})\boldsymbol{D}_1^{-1}(\boldsymbol{\tau})\right).$$

In order to use the ERFE estimator to make inference, an estimator of its VC-matrix is presented in **Theorem 3.3.3**. This will make it possible to construct large sample confidence intervals or hypothesis tests. The proposed VC-matrix estimator is robust and consistent, and is a generalization of the commonly advocated VC-matrix estimator proposed by White (1980).

Theorem 3.3.3.

Let

$$\widehat{\boldsymbol{D}}_0(\boldsymbol{\tau}) = \frac{1}{nm}(\boldsymbol{V}\otimes\widehat{\boldsymbol{X}^*})^{T}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}})\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}\boldsymbol{\varepsilon}^{*T}_{\boldsymbol{\tau}}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}})(\boldsymbol{V}\otimes\widehat{\boldsymbol{X}^*}),$$

$$\widehat{\boldsymbol{D}}_1(\boldsymbol{\tau}) = \frac{1}{nm}(\mathbf{I}_q\otimes\widehat{\boldsymbol{X}^*})^{T}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}})(\boldsymbol{V}\otimes\widehat{\boldsymbol{X}^*})$$

where the transformed data is obtained by pre-multiplying the initial data with the projection matrix $\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\boldsymbol{\tau}) = \mathbb{I}_{nmq} - \widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\boldsymbol{\tau})$ *and*

$$\widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\boldsymbol{\tau}) = (\boldsymbol{v} \otimes \boldsymbol{Z})\Big[(\boldsymbol{v} \otimes \boldsymbol{Z})^{T}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\tau}}})(\mathbb{1}_{q} \otimes \boldsymbol{Z})\Big]^{-1}(\mathbb{1}_{q} \otimes \boldsymbol{Z})^{T}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\tau}}}).$$

Then, for every fixed $\boldsymbol{\tau}$ *we have*

$$\widehat{\boldsymbol{D}}_{1}^{-1}(\boldsymbol{\tau})\widehat{\boldsymbol{D}}_{0}(\boldsymbol{\tau})\widehat{\boldsymbol{D}}_{1}^{-1}(\boldsymbol{\tau}) \xrightarrow{p} \boldsymbol{D}_{1}^{-1}(\boldsymbol{\tau})\boldsymbol{D}_{0}(\boldsymbol{\tau})\boldsymbol{D}_{1}^{-1}(\boldsymbol{\tau}).$$

We end this section with the result for a single τ.

Corollary 3.3.3.1.

Let

$$\widehat{\boldsymbol{D}}_{0}(\tau) = \frac{1}{nm}\sum_{i=1}^{n}\widehat{\boldsymbol{X}}_{i}^{*T}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^{*}_{i\tau}})\widehat{\boldsymbol{\varepsilon}^{*}_{i\tau}}\widehat{\boldsymbol{\varepsilon}^{*}_{i\tau}}^{T}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^{*}_{i\tau}})\widehat{\boldsymbol{X}}_{i}^{*},$$

$$\widehat{\boldsymbol{D}}_{1}(\tau) = \frac{1}{nm}\sum_{i=1}^{n}\widehat{\boldsymbol{X}}_{i}^{*T}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^{*}_{i\tau}})\widehat{\boldsymbol{X}}_{i}^{*}$$

with the corresponding projection matrices

$$\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau) = \mathbb{I}_{N} - \widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau) \text{ and } \widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau) = \boldsymbol{Z}(\boldsymbol{Z}^{T}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^{*}_{\tau}})\boldsymbol{Z})^{-1}\boldsymbol{Z}^{T}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^{*}_{\tau}}).$$

Then, under the above conditions and for every fixed τ*, we have*

$$\widehat{\boldsymbol{D}}_{1}^{-1}(\tau)\widehat{\boldsymbol{D}}_{0}(\tau)\widehat{\boldsymbol{D}}_{1}^{-1}(\tau) \xrightarrow{p} \boldsymbol{D}_{1}^{-1}(\tau)\boldsymbol{D}_{0}(\tau)\boldsymbol{D}_{1}^{-1}(\tau).$$

3.4 Simulations

In this section, the small sample performance of the ERFE estimators is evaluated through extensive simulation studies. The simulation design proposed here is replicated from Koenker (2004). The random samples are generated from the following

linear model

$$y_{ij} = \beta_0 + x_{ij}\beta_1 + \alpha_i + (1 + \gamma x_{ij})\varepsilon_{ij},\ 1 \leq i \leq n,\ 1 \leq j \leq m. \qquad (3.21)$$

Two versions of model (3.21) are considered with respect to the parameter $\gamma \in \{0, 1/10\}$. A location shift model (M_0) corresponding to $\gamma = 0$, which helps assess the performance of the estimators for an homoscedastic scenario and a location-scale shift model ($M_{1/10}$) corresponding to $\gamma = 1/10$ to assess the performance of the estimators in the presence of heteroscedasticity.

The corresponding ERFE model of M_0 is $\mu_\tau(y_{ij}) = \beta_{0\tau} + x_{ij}\beta_1 + \alpha_i$ where $\beta_{0\tau} = \beta_0 + \mu_\tau(\varepsilon_{ij})$, so that only the intercept term varies with τ and the expectile functions are parallel lines. The ERFE model related to the $M_{1/10}$ model is $\mu_\tau(y_{ij}) = \beta_{0\tau} + x_{ij}\beta_{1\tau} + \alpha_i$ where $\beta_{0\tau} = \beta_0 + \mu_\tau(\varepsilon_{ij})$ and $\beta_{1\tau} = \beta_1 + \gamma\mu_\tau(\varepsilon_{ij})$. Therefore, in the presence of heteroscedasticity both the intercept and the slope vary with τ.

The individual-specific effect α_i and the disturbance ε_{ij} are generated by the same distribution in three different models : normal distribution $\mathcal{N}(0, 1)$, Student distribution t_3 with 3 degrees of freedom, and chi-squared distribution χ^2_3 with 3 degrees of freedom. In addition, for each distribution the random error is centred on its τ-th expectile, $\mu_\tau(\varepsilon_{ij})$. The continuous explicative variable x is generated by a normal distribution in the location shift model and by a chi-squared distribution with 3 degrees of freedom in the location scale shift model. The parameters $\beta_0 = \beta_1 = 0$ are set to zero. The extensive simulation is carried out with 400 replications. In each case the focus is on the effect of the covariate (x_{ij}) at the expectiles $\tau \in \{0.25, 0.5, 0.75\}$. Following the various models and sample sizes, $n \times m \in$
$\{50, 100, 250\} \times \{5, 15, 25\}$, 54 different random samples are created. The results of

the ERFE regression are analyzed using the quantile regression with fixed-effects (QRFE) model of (Koenker, 2004) as a benchmark. The simulation of the QRFE model is carried out by choosing, for each distribution, the asymmetric points for which the quantiles are equal to the expectiles. For example, the Gaussian quantiles of $\tau = (0.33, 0.5, 0.67)$ correspond to the Gaussian expectiles of $\tau = (0.25, 0.5, 0.75)$. The quality of the estimators is evaluated by the distribution of their bias represented as box-plots. The standard deviation (SD) of the 400 parameter estimates is used as a benchmark to evaluate the average asymptotic standard errors (SE). We also reported the bias and the percentage of probability coverage ($P_{0.95}$) based on the sandwich estimate of the covariance matrix. All computations are performed with the R v3.4.0 statistical programming language (R Core Team, 2018). The function used to estimate the ERFE model is available on GitHub (https ://github.com/AmBarry/erfe). The QRFE model is estimated with the rqpd v0.6 package (Koenker et Bache, 2014) for R.

Figure 3.1 and **Figure 3.2** present the distribution of the bias of the ERFE and QRFE estimators respectively in the location-shift (M_0) and location-scale-shift scenario ($M_{1/10}$), when the error is normally distributed. The results for the other distributions (Student and Chi-Square) are in the **Supplementary material I**. Indicated in red are the results of the ERFE estimator and in blue the results of QRFE estimator. Columns present the results according to the values of $\tau \in (\tau_1, \tau_2, \tau_3)$; rows present the results according to the values of $m \in (5, 15, 25)$. Finally, on the x-axis we have the values of $n \in (50, 100, 250)$.

We observe that the bias is relatively small for both estimators, ERFE and QRFE, in the location-shift scenario, **Figure 3.1**. The distribution of the bias is symmetric

and centered around the true value, zero. The size of the interquartile range of the different box-plots as well as the number of extreme values decrease as the number of individuals and the number of within-observations per individuals increase, **Figure 3.1**. This is in line with the asymptotic properties of the estimators and the fact that the fixed-effects is consistent when the number of within-observation goes to infinity (Hansen, 2007). The ERFE estimator does better than the QRFE estimator whose box-plots are less centered with a larger range. This finding is valid depending on the type of distribution (symmetric or heavy tail) and the value of the asymmetric point (τ_1, τ_2, τ_3), see the **Supplementary material I**.

Naturally, there is more variability in the distribution of the bias in the location-scale-shift scenario ($M_{1/10}$), **Figure 3.2**. However, both estimators perform relatively well. The QRFE estimator does better than the ERFE in the location-scale shift scenario, **Figure 3.2**. Which is not surprising, because the QR is known to be more robust.

Table 3.1 reports the standard deviation (SD) and the asymptotic standard error (SE) of both methods for the normal distribution, other results are in the **Supplementary material I**. Importantly, the SD values show the variability of the boxplots observed in the previous Figures, **Figure 3.1-3.2**. As noted early, this variability decreases as n increases.

In general, the values of SD and SE are relatively identical, confirming the consistency and robustness of the robust sandwich variance-covariance ERFE estimator. Significantly, the "asymptotic" SE of the QRFE estimator corresponds to the bootstrap estimator. The estimation of the SE of the QRFE requires estimating the unknown

density which is computationally intensive and causes some numerical computational problems (Yin et Cai, 2005).

The results on the percentage of probability coverage $P_{0.95}$ are presented in **Table 3.2** and are in agreement with the previous simulation results. The results also show that the percentage coverage is below the nominal level (95%) for the ERFE method when one is not at the center of the distribution in the location scale-shift scenario.

We end this section by comparing the run-times of the ERFE and QRFE algorithms. We adjusted both methods on a simulated dataset with 4 covariates, 3 asymmetric points $\tau \in (0.25, 0.5, 0.75)$ and 3 different values of $n \in (500, 1000, 5000)$ and $m \in (5, 15, 25)$. The result shows that the ERFE algorithm outperforms the QRFE algorithm in all cases, **Table 3.3**. The results for $n = 5000$ and $m = 25$ are not printed because the QRFE algorithm used more memory than available in our computer (x86_64-redhat-linux-gnu (64-bit) with 125G of RAM). The numerical experiments is carried out using the microbenchmark package (Mersmann, 2018) from (R Core Team, 2017).

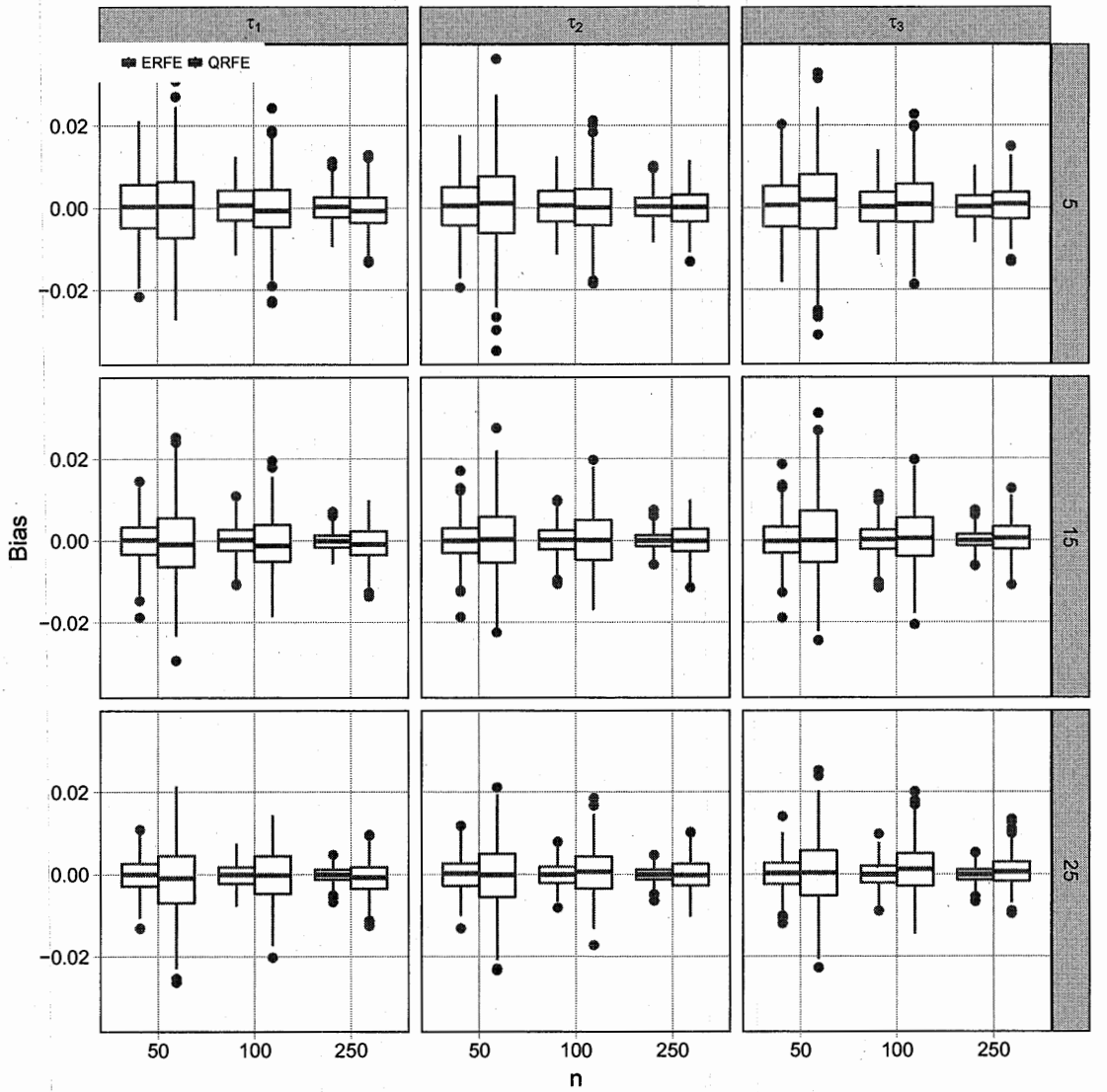

Figure 3.1: **Location-shift scenario** – Bias distribution (box-plot) of the ERFE and QRFE estimators according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim \mathcal{N}(0, 1)$.

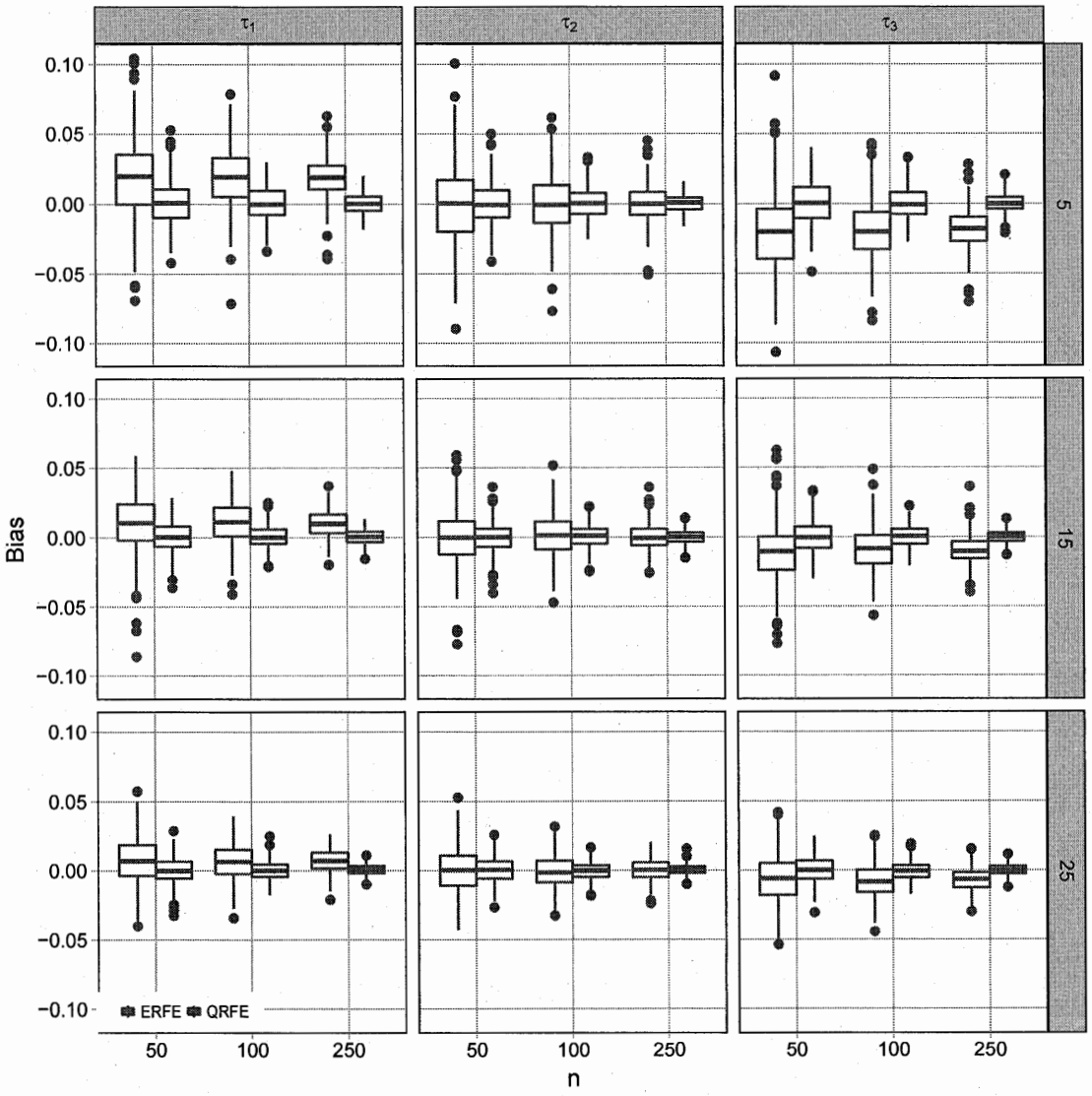

Figure 3.2: **Location-scale-shift scenario** – Bias distribution (box-plot) of the ERFE and QRFE estimators according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim \mathcal{N}(0, 1)$.

Tableau 3.1: Standard deviation (SD) and asymptotic standard error (SE) in the location-shift and location-scale-shift scenario with the error term $\varepsilon \sim \mathcal{N}(0,1)$.

		$m=5$				$m=10$			
		ERFE		QRFE		ERFE		QRFE	
	τ	SD	SE	SD	SE	SD	SE	SD	SE
		Location-shift model							
$n=50$	τ_1	0.0074	0.0058	0.0100	0.0105	0.0049	0.0044	0.0091	0.0088
	τ_2	0.0071	0.0065	0.0103	0.0103	0.0048	0.0044	0.0087	0.0086
	τ_3	0.0073	0.0072	0.0106	0.0106	0.0051	0.0048	0.0088	0.0087
$n=100$	τ_1	0.0048	0.0041	0.0071	0.0074	0.0037	0.0031	0.0066	0.0064
	τ_2	0.0047	0.0046	0.0069	0.0072	0.0035	0.0032	0.0065	0.0063
	τ_3	0.0049	0.0051	0.0072	0.0075	0.0036	0.0035	0.0066	0.0064
$n=250$	τ_1	0.0034	0.0026	0.0046	0.0047	0.0022	0.0020	0.0041	0.0041
	τ_2	0.0033	0.0029	0.0045	0.0046	0.0021	0.0020	0.0040	0.0040
	τ_3	0.0035	0.0033	0.0047	0.0047	0.0022	0.0022	0.0040	0.0041
		Location-scale-shift model							
$n=50$	τ_1	0.0282	0.0160	0.0153	0.0155	0.0208	0.0141	0.0109	0.0106
	τ_2	0.0278	0.0209	0.0150	0.0147	0.0198	0.0162	0.0108	0.0102
	τ_3	0.0278	0.0236	0.0154	0.0151	0.0201	0.0178	0.0112	0.0108
$n=100$	τ_1	0.0208	0.0126	0.0117	0.0113	0.0155	0.0110	0.0078	0.0077
	τ_2	0.0203	0.0164	0.0104	0.0106	0.0149	0.0127	0.0076	0.0074
	τ_3	0.0203	0.0186	0.0112	0.0109	0.0153	0.0140	0.0077	0.0077
$n=250$	τ_1	0.0134	0.0087	0.0073	0.0071	0.0098	0.0074	0.0052	0.0050
	τ_2	0.0131	0.0113	0.0067	0.0068	0.0095	0.0085	0.0049	0.0047
	τ_3	0.0132	0.0129	0.0069	0.0071	0.0100	0.0094	0.0048	0.0050

Tableau 3.2: Bias and probability coverage in the location-shift and location-scale-shift scenario with the error term $\varepsilon \sim \mathcal{N}(0,1)$.

		$m=5$				$m=10$			
		Bias		$P_{0.95}$		Bias		$P_{0.95}$	
	τ	ERFE	QRFE	ERFE	QRFE	ERFE	QRFE	ERFE	QRFE
					Location-shift model				
$n=50$	τ_1	0.000	0.000	0.978	0.982	0.000	0.000	0.990	0.972
	τ_2	0.000	0.001	0.995	0.975	0.000	0.000	0.992	0.988
	τ_3	0.000	0.001	0.992	0.980	0.000	0.001	0.992	0.982
$n=100$	τ_1	0.001	0.000	0.998	0.988	0.000	-0.001	1.000	1.000
	τ_2	0.000	0.000	1.000	0.985	0.000	0.000	1.000	0.992
	τ_3	0.000	0.001	1.000	0.985	0.000	0.001	1.000	0.995
$n=250$	τ_1	0.000	-0.001	0.998	1.000	0.000	-0.001	1.000	0.998
	τ_2	0.000	0.000	1.000	1.000	0.000	0.000	1.000	1.000
	τ_3	0.000	0.001	1.000	0.998	0.000	0.001	1.000	0.998
					Location-scale-shift model				
$n=50$	τ_1	0.019	0.001	0.722	0.975	0.010	0.001	0.840	0.970
	τ_2	0.000	0.000	0.910	0.958	0.000	0.000	0.940	0.965
	τ_3	-0.020	0.000	0.852	0.972	-0.010	0.000	0.915	0.978
$n=100$	τ_1	0.019	0.000	0.660	0.975	0.011	0.001	0.790	0.982
	τ_2	0.000	0.001	0.942	0.980	0.001	0.001	0.970	0.972
	τ_3	-0.019	0.000	0.838	0.970	-0.009	0.000	0.925	0.982
$n=250$	τ_1	0.019	0.000	0.588	0.988	0.010	0.000	0.818	0.995
	τ_2	0.000	0.000	0.980	0.998	0.000	0.000	0.982	0.998
	τ_3	-0.018	0.000	0.792	0.998	-0.010	0.000	0.895	0.998

Tableau 3.3: Comparison of the run-times (in milliseconds) of the ERFE and QRFE method

n	500			1000			5000		
m	5	15	25	5	15	25	5	15	25
ERFE	506	1412	3289	1205	4665	11470	16260	112567	-
QRFE	8704	24662	42323	16334	51512	89574	90982	321559	-

3.5 Application

Returns to schooling (also known as returns to education) estimated by the Mincer equation (Lemieux, 2006) is arguably the most widely studied topic in empirical economics. It is often presented in standard econometric textbooks (Baltagi, 2008; Greene, 2012; Cameron et Trivedi, 2008) as an example of an endogeneity model. Indeed, there is a potential correlation between individual ability and the other covariates such as education or experience. In the presence of endogeneity, the FE method is often preferred to other methods that estimate the average effect. Despite the fact that it does not estimate the effect of the time-invariant covariates, the FE estimator is consistent even if the subject-specific effects are correlated with the covariates of the model (Baltagi, 2008). We estimate the earnings equation using the ERFE estimator which inherits the favourable properties of the FE estimator. In addition to obtaining the results of the FE estimator on the average salary, available here (Baltagi, 2008; Baltagi et Khanti-Akom, 1990; Cornwell et Rupert, 1988), we will also capture the results of the FE estimator on the entire wage distribution. Consequently, the ERFE model allows for both endogeneity resulting from unmeasured

factors and heterogeneity in the effect of the covariates. For example, we'll be able to determine if being unionized is more advantageous for those who have a small or a large salary. We also consider the QRFE model in our analysis.

We replicated Baltagi and Khanti-Akon's study (Baltagi et Khanti-Akom, 1990) using the Panel Study of Income Dynamics (PSID) dataset (Cornwell et Rupert, 1988). The dataset is a cohort of 595 individuals observed over the period 1976–1982. The respondents, aged between 18 and 65 in 1976, are those who reported a positive wage in some private, non-farm employment for all 7 years, (Cornwell et Rupert, 1988).

The log wage is the dependent variable and is regressed on weeks worked (WKS), years of full-time work experience (EXP), occupation (OCC=1, if the individual is in a blue-collar occupation), residence (SOUTH = 1, SMSA = 1, if the individual resides in the South, or in a standard metropolitan statistical area), marital status (MS = 1, if the individual is married), industry (IND = 1, if the individual works in a manufacturing industry) and union coverage (UNION = 1, if the individual's wage is set by a union contract). The corresponding Mincer equation for both methods ERFE and QRFE is

$$\begin{aligned}\mu_\tau(\log(\text{Wage}_{ij})) \text{ or } q_\tau(\log(\text{Wage}_{ij})) = {} & \beta_{0\tau} + \beta_1 \text{WKS}_{ij} + \beta_2 \text{EXP}_{ij} + \beta_3 \text{EXP}_{ij}^2 \\ & + \beta_4 \text{UNION}_{ij} + \beta_5 \text{IND}_{ij} + \beta_6 \text{MS}_{ij} + \beta_7 \text{OCC}_{ij} \\ & + \beta_8 \text{SOUTH}_{ij} + \beta_9 \text{SMSA}_{ij} + \alpha_i.\end{aligned}$$

We estimated the conditional expectile and quantile of the log wage distribution using 91 asymmetric points $(0.05, 0.06, 0.07, \ldots, 0.95)$. We generated the confidence intervals using the asymptotic and the bootstrap standard error respectively for

the ERFE and QRFE model. The results of both models are presented in **Figure 3.5-3.6**. We also presented the parameter estimates, the confidence intervals and the p-values of some variables for five asymmetric point $(0.1, 0.25, 0.5, 0.75, 0.9)$ in **Table 3.4**. The complete results for all the variables are in the **Supplementary material I**.

To the best of our knowledge, there are few or no empirical studies in the literature comparing ER and QR as presented by Newey et Powell (1987) and Koenker et Bassett (1978). To fill this gap, we start by showing the results of the ER and QR models adjusted to the PSID dataset, **Figure 3.3-3.4**. Notice that, for a fixed τ, we cannot compare expectile of level τ and the corresponding quantile of level τ. Comparing the two approaches must be done globally.

The results presented in **Figure 3.3-3.4**, show the potential of both ER and QR models to reveal the heterogeneity of the effects of the covariates on the distribution of the response variable. It seems that the effects of certain covariates, like EXP or WKS for example, are constant according to the expectile level, **Figure 3.3-3.4**. This would mean that estimation of the covariate effects on the conditional mean of the response variable would be enough to analyze the data. Nevertheless, there are other variables (UNION, IND and SOUTH, for example) for which the heterogeneity of the effects is clearly visible.

Curves from the ER method are smoother than those from the QR method which seem to be more unstable. Despite that, both models show similar results with same

scale and trend. These results, **Figure 3.3-3.4**, show that the ER model and its extensions are reliable alternatives for the QR model whose computational cost grows with the increase of the sample size. We also received warnings of non-uniqueness of the solutions from the quantreg package (Koenker, 2018).

Figure 3.5 and **Figure 3.6** show results of the ERFE and QRFE models. Again, the results show that, we have to go beyond the mean or median effect in order to capture the heterogeneity of the effects. The FE model, like other methods that estimate the mean effect, is not sufficient to analyze the returns to schooling because the impact of most of the covariates vary across the wage distribution. For example, the ERFE and the QRFE results indicate that being part of a union is significantly more advantageous for low wages than for high wages. This result, on the effects of unions on the structure of wages, is consistent with the results obtained by Card (1996). We see also that the effect of the covariates is not significant in every side of the response distribution. For example, the ERFE effect of the OCC covariate is not significant on the interquartile range of the response variable but the effect is significant on the tail of the response distribution. In other words, the observed salary difference between withe/blue collar is statistically significant (at 10%) only in the group with low/high wages, **Table 3.4**. That is, there is no significant difference between withe/blue collar with average salary. In the other hand, the QRFE effect of the OCC covariate is significant in the center and right of the response variable. Similar conclusions can be drawn for the other covariates : the rate of increase/decrease in salary is not the same depending on whether one makes a small, medium or large salary.

After crossing the computational hurdle, we observe that the course of the effects of the QRFE model, according to the values of τ, is more volatile and less smooth than

the ERFE model, **Figure 3.5-3.6**. It is more difficult to identify an overall trend of the heterogeneity of the effects with the QRFE model. For example, in **Table 3.4** the QRFE effect of the IND covariate is decreasing between the asymmetric points, 0.1 and 0.25 and then increasing between 0.25 and 0.9.

We observe that the QRFE results display higher rates and different trends in comparison to the ERFE model for some covariates. In general, the range of the QRFE estimates is 10 times higher than that of the ERFE estimates, whose values are almost 0, **Figure 3.5-3.6**. This difference is not surprising when we consider the distribution of these covariates and the characteristics of the ERFE estimator. On the one hand, these covariates, including marital status (MS), have very small within-subject variation and are therefore practically time-invariant covariates, **Figure 3.7**. On the other hand, the ERFE estimator is a within estimator, whose expression depends on the within variation of the covariates. In the light of the above explanation, we can expect these results.

There are alternatives in the literature (Cornwell et Rupert, 1988; Baltagi et Khanti-Akom, 1990) that have been proposed to circumvent this limit (lack of inference for the time invariant covariates), while keeping the qualities of the FE model. Future research should investigate the possibility of adapting these methods (Cornwell et Rupert, 1988; Baltagi et Khanti-Akom, 1990) under the expectile regression framework.

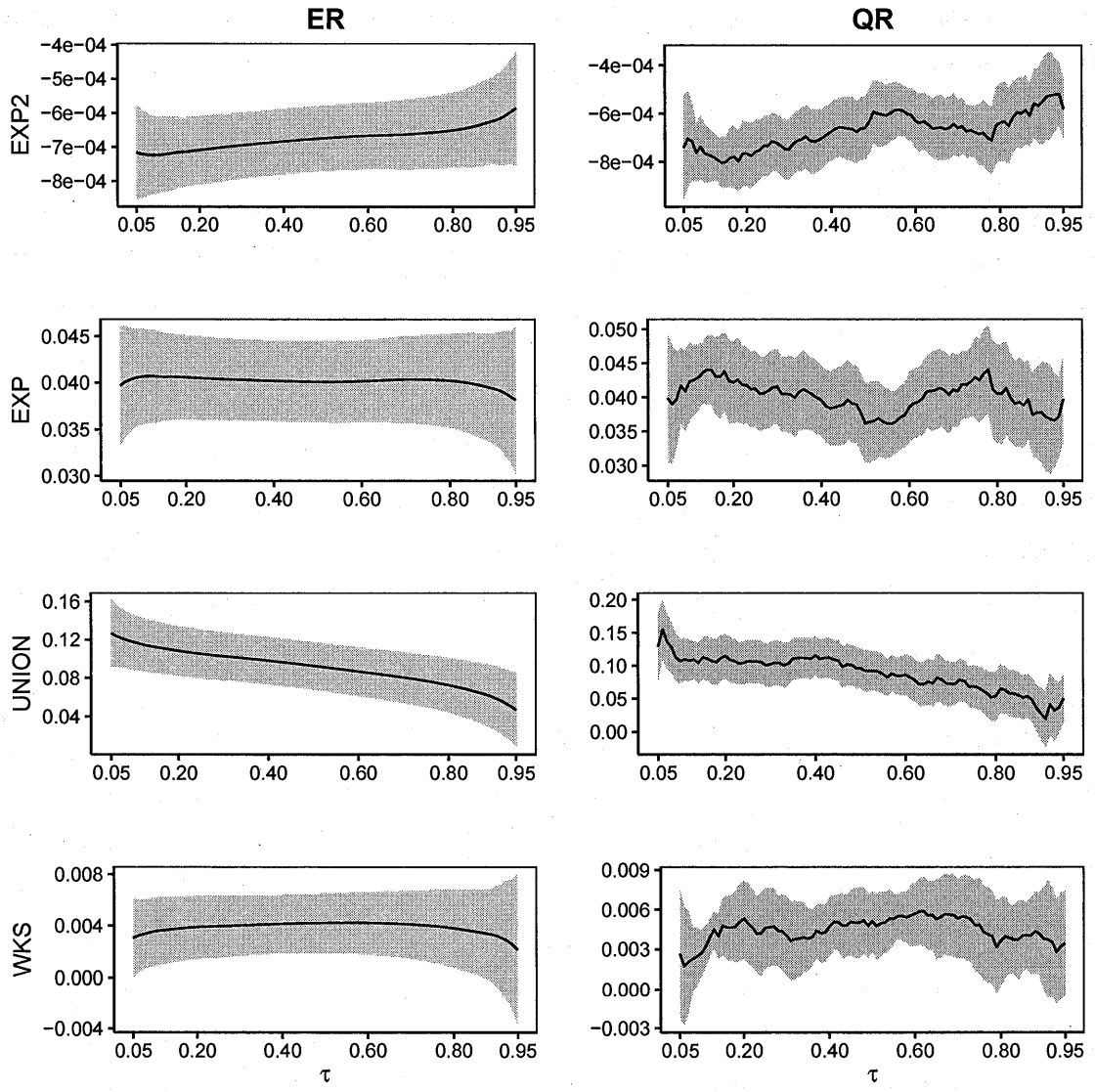

Figure 3.3: ER and QR estimated coefficients with estimated confidence intervals, Part I.

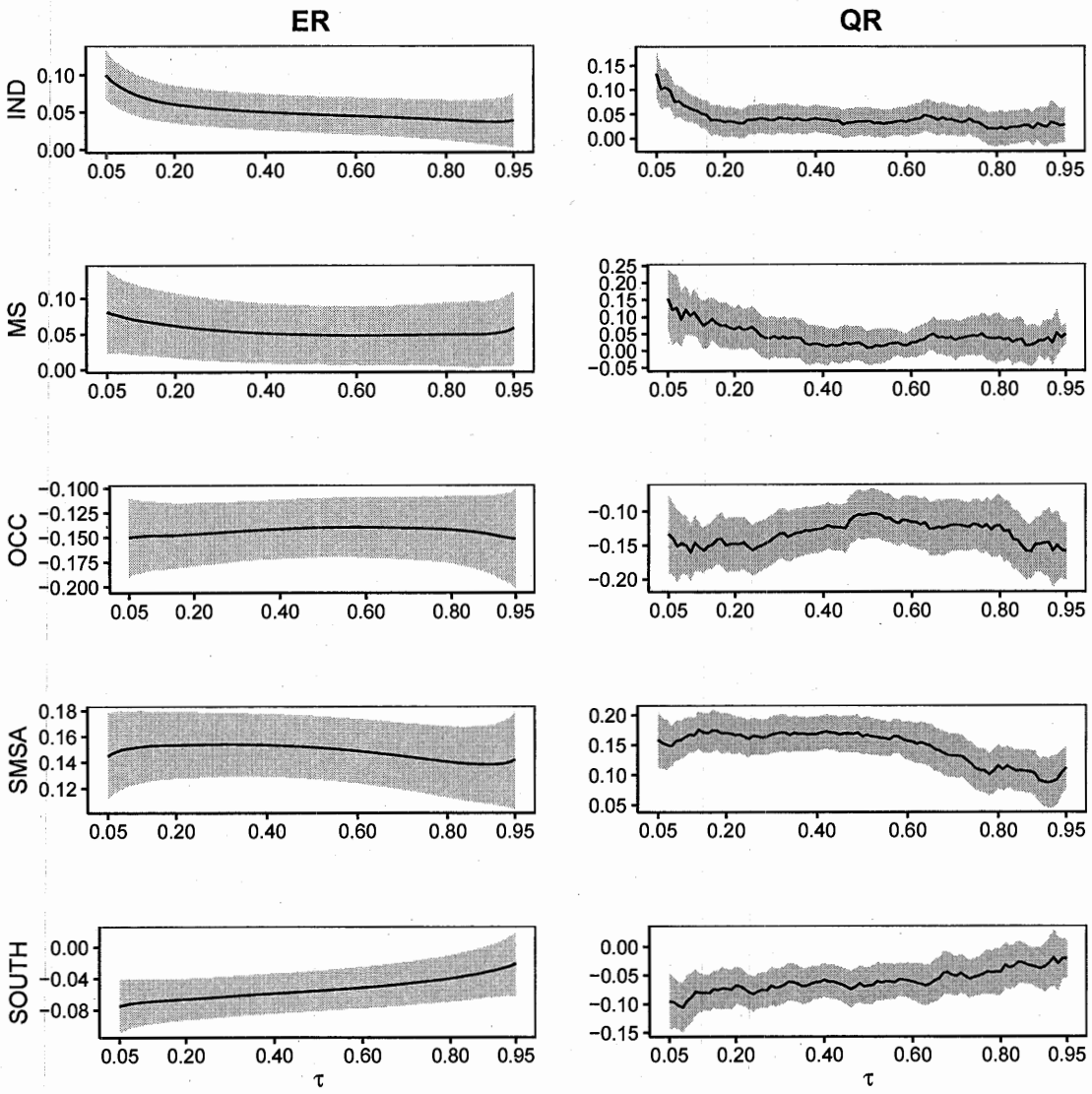

Figure 3.4: ER and QR estimated coefficients with estimated confidence intervals, Part II.

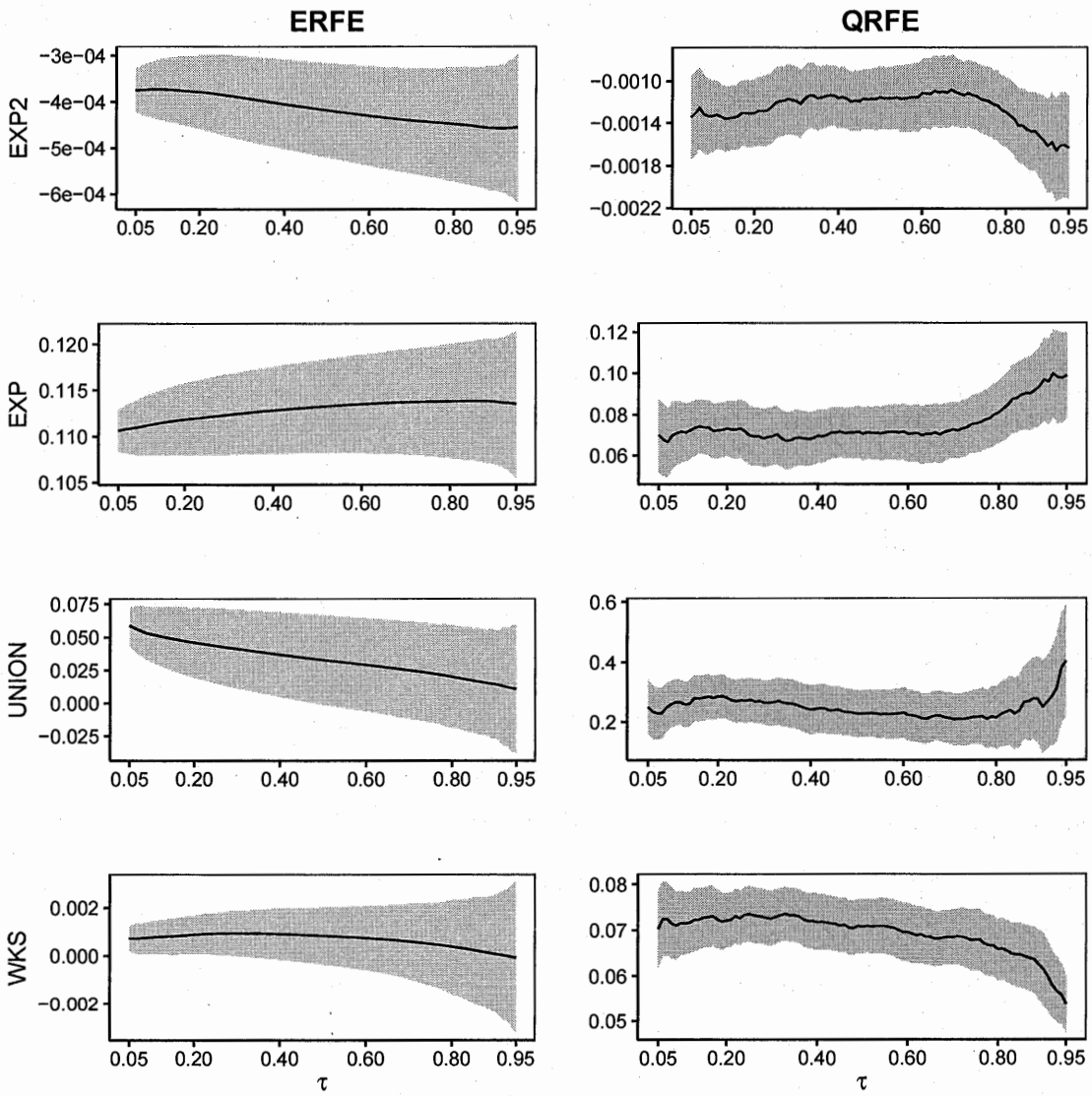

Figure 3.5: ERFE and QRFE estimated coefficients with estimated confidence intervals, Part I.

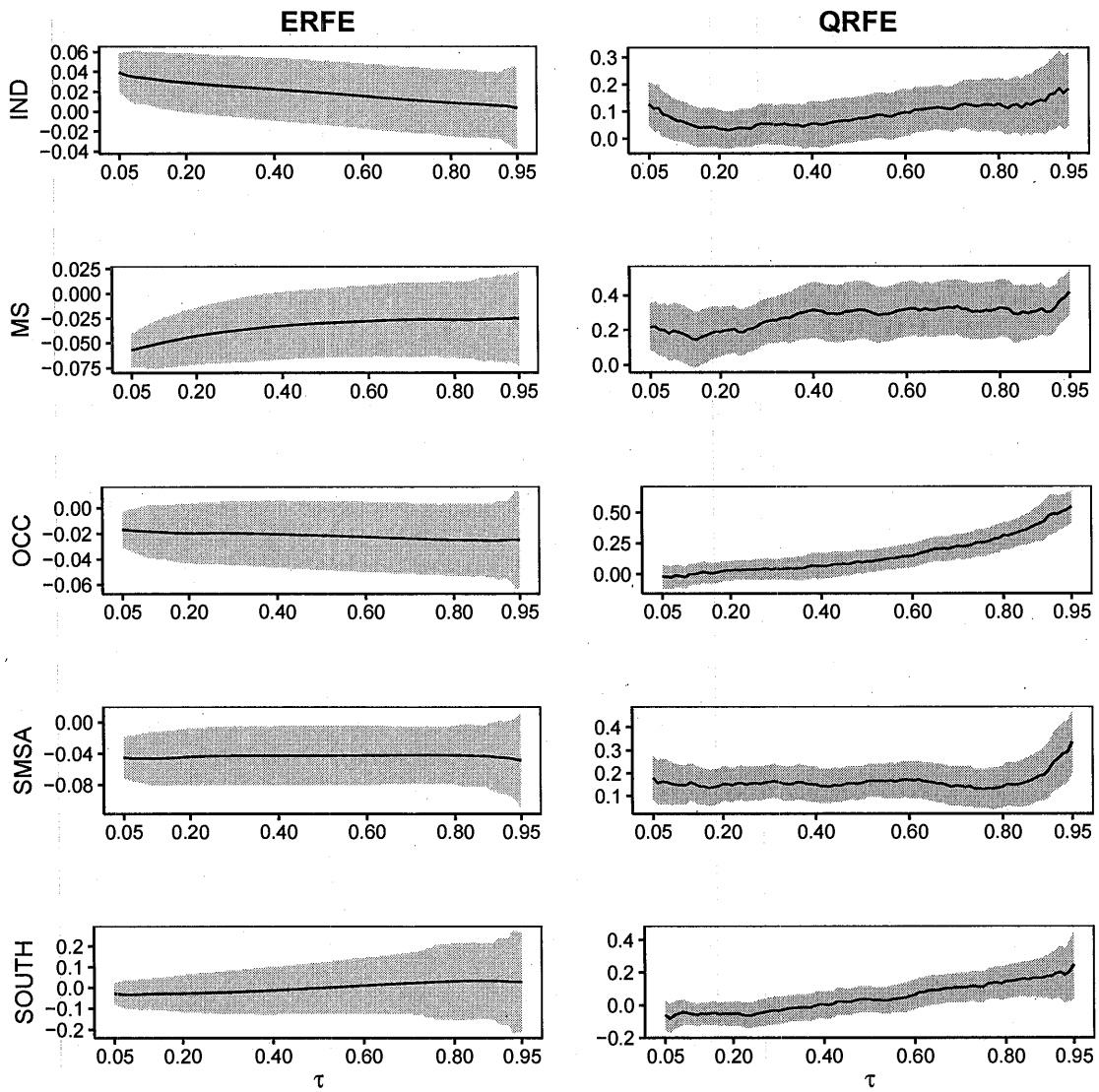

Figure 3.6: ERFE and QRFE estimated coefficients with estimated confidence intervals, Part II.

Tableau 3.4: Parameters estimates (Est) and p-values obtained from the ERFE and QRFE methods at five percentiles, $\tau = (0.1, 0.25, 0.5, 0.75, 0.9)$.

Var	τ	0.1	0.25	0.5	0.75	0.9
IND	ERFE	0.0340***	0.0269*	0.0192	0.0104	0.0063
	QRFE	0.0746**	0.0396	0.0745*	0.1205***	0.1391**
OCC	ERFE	-0.0179*	-0.0195	-0.0215	-0.0246*	-0.0255*
	QRFE	-0.0238	0.0368	0.1042**	0.2543***	0.4722***
UNION	ERFE	0.0524***	0.0435***	0.0328**	0.0228	0.0144
	QRFE	0.2589***	0.2724***	0.2271***	0.2158***	0.2497***

***significant at 1% level, **significant at 5% level, *significant at 10% level.

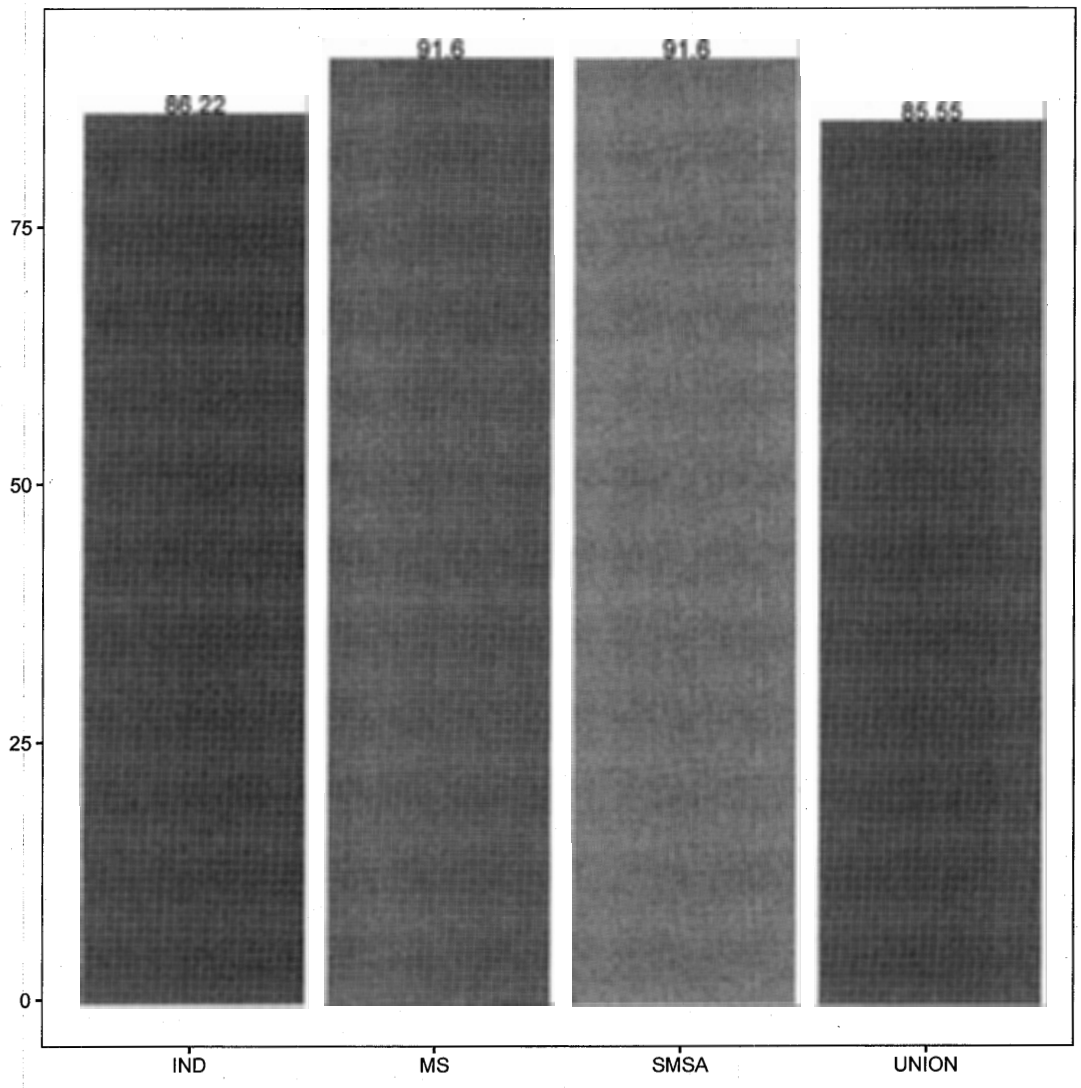

Figure 3.7: Percentage of subject with constant within-subject observations.

3.6 Conclusion

We introduced the ERFE model which follows from the combination of the weighted asymmetric least squares regression (ER) and the FE model. The ERFE model inherits the attractive properties of both models. As with the FE model, the ERFE model is an instrumental variable model that takes into account the possible correlation between covariates of the model and those missing from the model. In addition, the ERFE model allows us to study the influence of covariates on the location, scale and shape of the conditional distribution of the response.

We show that the ERFE estimator is an iterative-within-transformation estimator. That is, the ERFE estimator can be derived by using iteratively the within-transformation strategy to concentrate out the incidental parameter from the model. We derive its asymptotic properties and suggest an estimator of its variance covariance matrix. The exhaustive simulations show that the ERFE estimator outperforms the QRFE estimator in the location-shift scenario. In the location-scale-shift scenario, the QRFE estimator outperforms the ERFE estimator. This result is not surprising because QR is known to be more robust than ER which has similar properties to OLS. Application of regression diagnostic can mitigate the influence of outliers and enhance the accuracy of ER estimators. Finally, we illustrate the method using the Panel Study of Income Dynamics (PSID) dataset from Baltagi et Khanti-Akom (1990).

It is known that QR is much more robust than ER. However, ER can be very useful to the statistical field. ER is computationally efficient compared to QR which is an attractive property in the high dimensional framework. In addition, ER naturally

extend classical statistical models like the generalized estimating equation or the within estimator, as shown in this work, which is not always possible for the QR. In regard to all of this, we think that ER have its place between the classical statistical models and the QR models.

The ERFE model enhances the FE model, which corresponds to the ERFE model of level $\tau = 0.5$. At the same time, it suffers from the same limitation. That is, all time-invariant covariates are eliminated by data transformation, so their coefficients cannot be estimated. Future research should investigate alternatives to circumvent this limit. Application of penalties on the individual fixed-effects parameter is a promising avenue.

3.7 Appendix

3.7.1 Supplementary Material I : Additional results

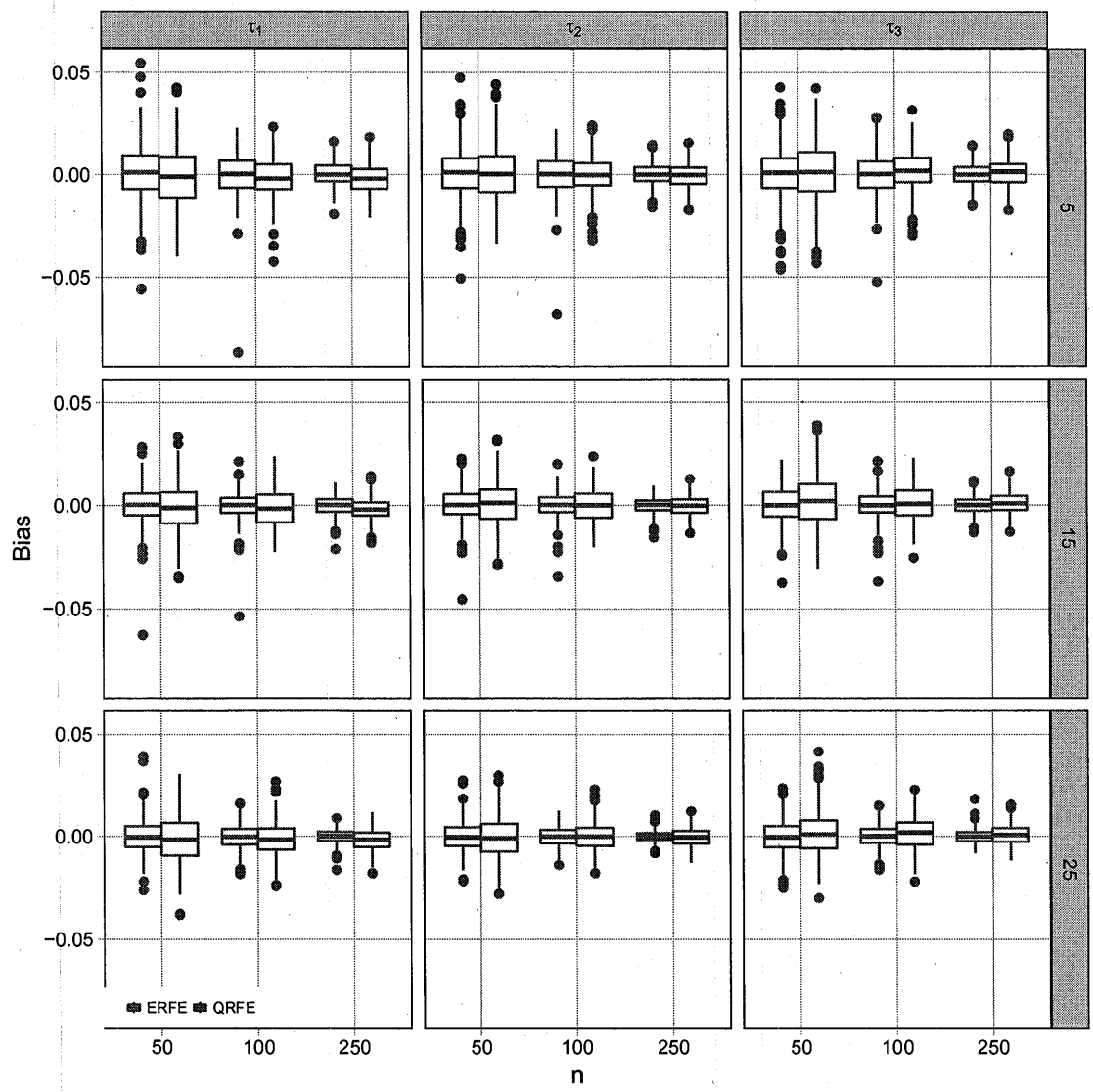

Figure 3.8: **Location-shift scenario** – Bias distribution (box-plot) of the ERFE and QRFE estimators according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim t_3$.

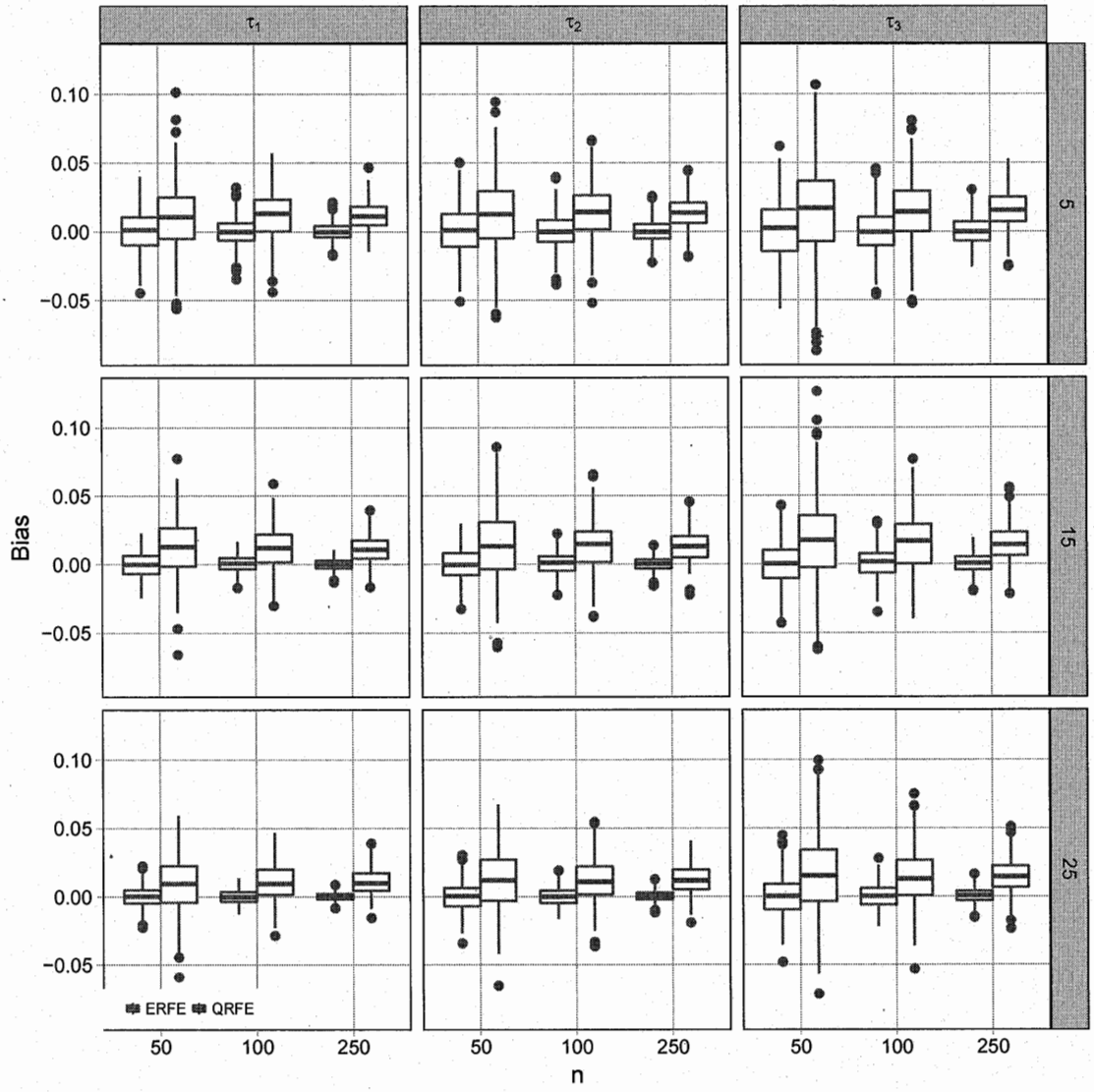

Figure 3.9: **Location-shift scenario** – Bias distribution (box-plot) of the ERFE and QRFE estimators according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim \chi^3_2$.

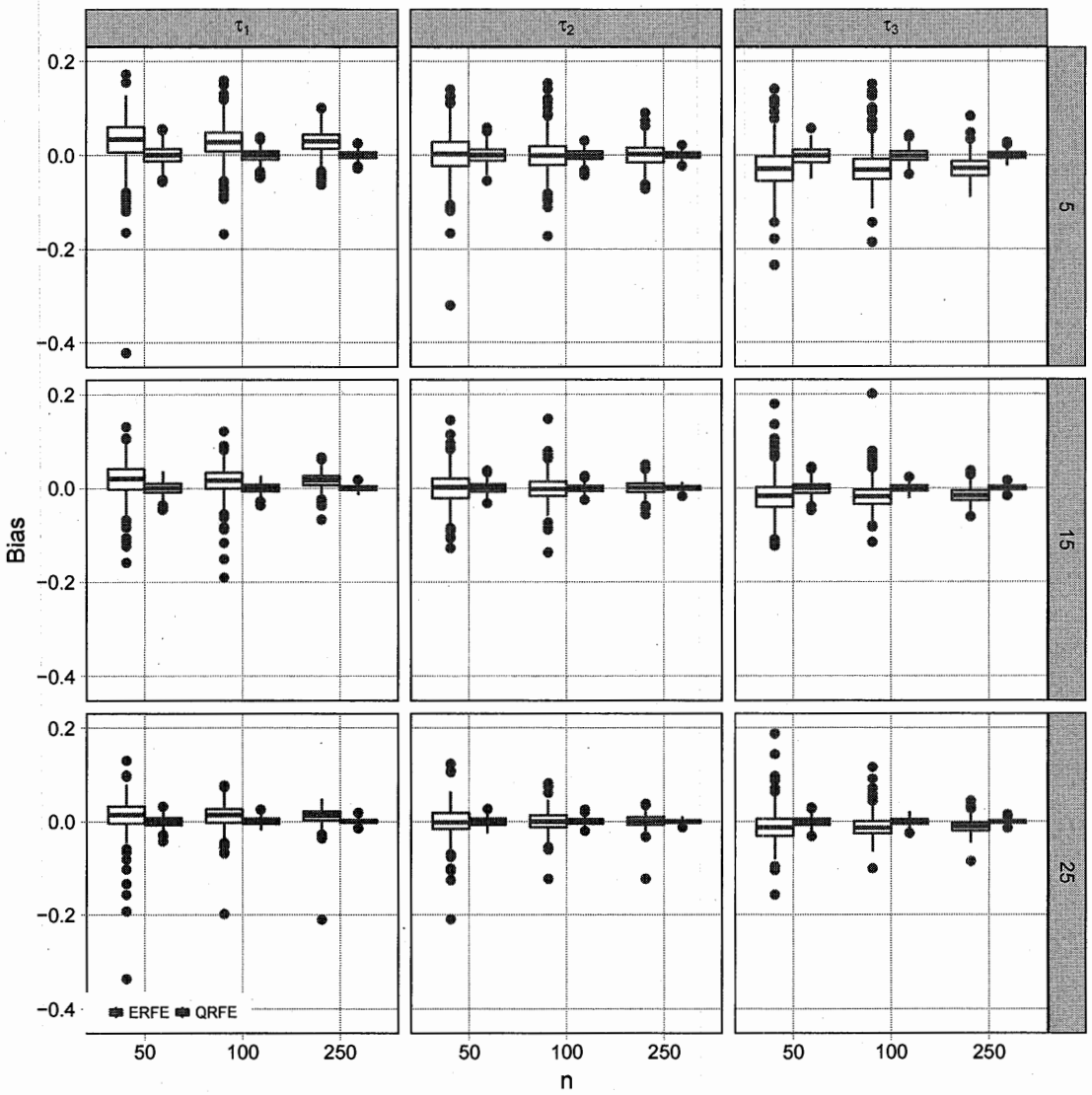

Figure 3.10: **Location-scale-shift scenario** – Bias distribution (box-plot) of the ERFE and QRFE estimators according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim t_3$.

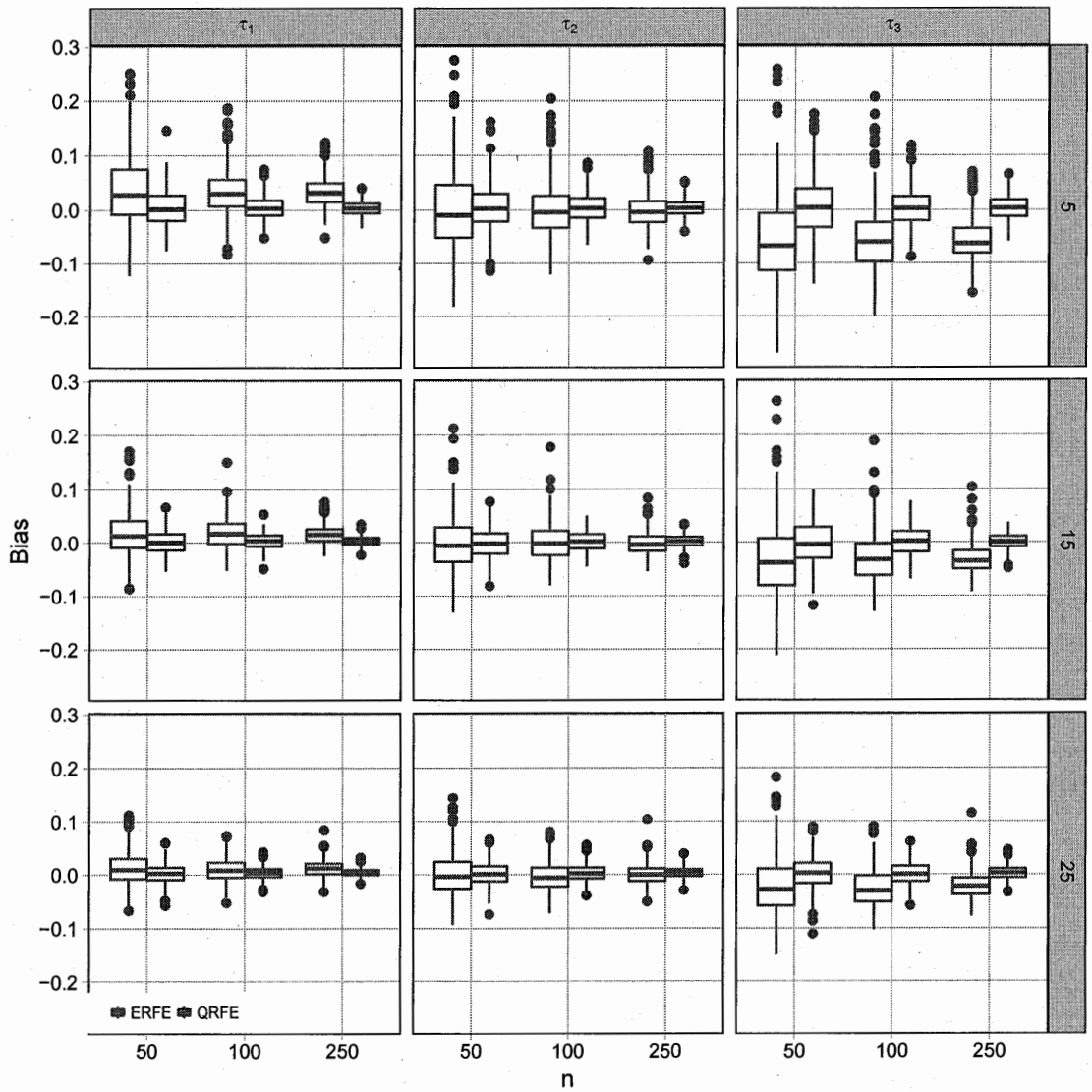

Figure 3.11: **Location-scale-shift scenario** – Bias distribution (box-plot) of the ERFE and QRFE estimators according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim \chi^3_2$.

Tableau 3.5: Standard deviation (SD) and asymptotic standard error (SE) of the location-shift and location-scale-shift scenario, $n = 50$.

		$n = 50,\ m = 5$				$n = 50,\ m = 10$			
		ERFE		QRFE		ERFE		QRFE	
	τ	SD	SE	SD	SE	SD	SE	SD	SE
					Location-shift model				
$\mathcal{N}$	τ_1	0.0074	0.0058	0.0100	0.0105	0.0049	0.0044	0.0091	0.0088
	τ_2	0.0071	0.0065	0.0103	0.0103	0.0048	0.0044	0.0087	0.0086
	τ_3	0.0073	0.0072	0.0106	0.0106	0.0051	0.0048	0.0088	0.0087
$\mathcal{T}$	τ_1	0.0135	0.0098	0.0143	0.0140	0.0093	0.0080	0.0116	0.0121
	τ_2	0.0124	0.0109	0.0133	0.0129	0.0079	0.0076	0.0108	0.0111
	τ_3	0.0129	0.0125	0.0142	0.0143	0.0087	0.0088	0.0124	0.0121
χ^2_3	τ_1	0.0143	0.0112	0.0238	0.0242	0.0088	0.0079	0.0207	0.0213
	τ_2	0.0173	0.0158	0.0270	0.0283	0.0117	0.0110	0.0237	0.0245
	τ_3	0.0214	0.0212	0.0328	0.0328	0.0160	0.0153	0.0289	0.0284
					Location-scale-shift model				
$\mathcal{N}$	τ_1	0.0282	0.0160	0.0153	0.0155	0.0208	0.0141	0.0109	0.0106
	τ_2	0.0278	0.0209	0.0150	0.0147	0.0198	0.0162	0.0108	0.0102
	τ_3	0.0278	0.0236	0.0154	0.0151	0.0201	0.0178	0.0112	0.0108
$\mathcal{T}$	τ_1	0.0517	0.0254	0.0199	0.0195	0.0372	0.0234	0.0137	0.0132
	τ_2	0.0472	0.0327	0.0178	0.0162	0.0335	0.0260	0.0118	0.0113
	τ_3	0.0473	0.0374	0.0189	0.0189	0.0361	0.0296	0.0135	0.0131
χ^2_3	τ_1	0.0614	0.0325	0.0328	0.0292	0.0409	0.0268	0.0226	0.0202
	τ_2	0.0708	0.0493	0.0413	0.0381	0.0515	0.0392	0.0285	0.0259
	τ_3	0.0812	0.0641	0.0543	0.0494	0.0665	0.0525	0.0377	0.0343

Tableau 3.6: Standard deviation (SD) and asymptotic standard error (SE) of the location-shift and location-scale-shift scenario, $n = 100$.

		$n = 100,\ m = 5$				$n = 100,\ m = 10$			
		ERFE		QRFE		ERFE		QRFE	
	τ	SD	SE	SD	SE	SD	SE	SD	SE
					Location-shift model				
$\mathcal{N}$	τ_1	0.0048	0.0041	0.0071	0.0074	0.0037	0.0031	0.0066	0.0064
	τ_2	0.0047	0.0046	0.0069	0.0072	0.0035	0.0032	0.0065	0.0063
	τ_3	0.0049	0.0051	0.0072	0.0075	0.0036	0.0035	0.0066	0.0064
$\mathcal{T}$	τ_1	0.0100	0.0072	0.0096	0.0100	0.0066	0.0058	0.0093	0.0085
	τ_2	0.0093	0.0080	0.0087	0.0093	0.0058	0.0055	0.0080	0.0078
	τ_3	0.0097	0.0091	0.0096	0.0100	0.0065	0.0063	0.0085	0.0085
χ^2_3	τ_1	0.0101	0.0080	0.0176	0.0173	0.0061	0.0056	0.0150	0.0151
	τ_2	0.0125	0.0113	0.0199	0.0200	0.0080	0.0078	0.0172	0.0175
	τ_3	0.0154	0.0151	0.0231	0.0234	0.0109	0.0108	0.0213	0.0204
					Location-shift-scale model				
$\mathcal{N}$	τ_1	0.0208	0.0126	0.0117	0.0113	0.0155	0.0110	0.0078	0.0077
	τ_2	0.0203	0.0164	0.0104	0.0106	0.0149	0.0127	0.0076	0.0074
	τ_3	0.0203	0.0186	0.0112	0.0109	0.0153	0.0140	0.0077	0.0077
$\mathcal{T}$	τ_1	0.0368	0.0204	0.0141	0.0139	0.0308	0.0195	0.0100	0.0095
	τ_2	0.0369	0.0264	0.0126	0.0119	0.0261	0.0208	0.0083	0.0081
	τ_3	0.0400	0.0307	0.0139	0.0135	0.0277	0.0237	0.0089	0.0096
χ^2_3	τ_1	0.0413	0.0255	0.0213	0.0212	0.0287	0.0213	0.0149	0.0150
	τ_2	0.0495	0.0389	0.0265	0.0270	0.0367	0.0306	0.0188	0.0190
	τ_3	0.0596	0.0508	0.0348	0.0354	0.0474	0.0412	0.0260	0.0256

Tableau 3.7: Standard deviation (SD) and asymptotic standard error (SE) of the location-shift and location-scale-shift scenario, $n = 250$.

		$n = 250,\ m = 5$				$n = 250,\ m = 10$			
		ERFE		QRFE		ERFE		QRFE	
	τ	SD	SE	SD	SE	SD	SE	SD	SE
					Location-shift model				
$\mathcal{N}$	τ_1	0.0034	0.0026	0.0046	0.0047	0.0022	0.0020	0.0041	0.0041
	τ_2	0.0033	0.0029	0.0045	0.0046	0.0021	0.0020	0.0040	0.0040
	τ_3	0.0035	0.0033	0.0047	0.0047	0.0022	0.0022	0.0040	0.0041
$\mathcal{T}$	τ_1	0.0054	0.0045	0.0065	0.0063	0.0042	0.0037	0.0051	0.0053
	τ_2	0.0049	0.0049	0.0057	0.0058	0.0037	0.0035	0.0047	0.0048
	τ_3	0.0051	0.0056	0.0063	0.0064	0.0041	0.0041	0.0055	0.0053
χ^2_3	τ_1	0.0062	0.0050	0.0104	0.0108	0.0038	0.0036	0.0091	0.0094
	τ_2	0.0077	0.0071	0.0117	0.0124	0.0050	0.0050	0.0106	0.0110
	τ_3	0.0098	0.0096	0.0135	0.0147	0.0068	0.0069	0.0132	0.0129
					Location-shift-scale model				
$\mathcal{N}$	τ_1	0.0134	0.0087	0.0073	0.0071	0.0098	0.0074	0.0052	0.0050
	τ_2	0.0131	0.0113	0.0067	0.0068	0.0095	0.0085	0.0049	0.0047
	τ_3	0.0132	0.0129	0.0069	0.0071	0.0100	0.0094	0.0048	0.0050
$\mathcal{T}$	τ_1	0.0238	0.0139	0.0086	0.0085	0.0167	0.0124	0.0057	0.0060
	τ_2	0.0228	0.0176	0.0076	0.0074	0.0148	0.0134	0.0052	0.0052
	τ_3	0.0234	0.0203	0.0089	0.0087	0.0160	0.0155	0.0062	0.0062
χ^2_3	τ_1	0.0284	0.0172	0.0131	0.0135	0.0165	0.0138	0.0091	0.0094
	τ_2	0.0329	0.0264	0.0165	0.0171	0.0209	0.0202	0.0115	0.0121
	τ_3	0.0388	0.0346	0.0231	0.0231	0.0271	0.0276	0.0145	0.0159

Tableau 3.8: Bias and probability coverage in the location-shift and location-scale-shift scenario with the error term $\varepsilon \sim t_3$.

		$m = 5$				$m = 10$			
		Bias		$P_{0.95}$		Bias		$P_{0.95}$	
	τ	ERFE	QRFE	ERFE	QRFE	ERFE	QRFE	ERFE	QRFE
					Location-shift model				
$n = 50$	τ_1	0.001	-0.001	0.940	0.962	0.000	-0.001	0.978	0.985
	τ_2	0.001	0.000	0.970	0.972	0.000	0.000	0.985	0.985
	τ_3	0.001	0.001	0.980	0.970	0.000	0.001	0.990	0.985
$n = 100$	τ_1	0.000	-0.001	0.982	0.990	0.000	-0.002	0.995	0.990
	τ_2	0.000	0.000	0.990	0.988	0.000	0.000	0.995	0.995
	τ_3	0.000	0.002	0.992	0.975	0.000	0.001	0.995	0.992
$n = 250$	τ_1	0.000	-0.002	0.985	0.995	0.000	-0.002	1.000	0.995
	τ_2	0.000	0.000	0.995	1.000	0.000	0.000	1.000	0.998
	τ_3	0.000	0.001	0.998	0.982	0.000	0.001	1.000	0.998
					Location-scale-shift model				
$n = 50$	τ_1	0.029	0.000	0.628	0.952	0.017	0.001	0.740	0.978
	τ_2	0.001	0.001	0.890	0.948	0.000	0.000	0.945	0.975
	τ_3	-0.028	-0.002	0.802	0.955	-0.016	-0.001	0.848	0.972
$n = 100$	τ_1	0.028	-0.001	0.612	0.972	0.014	0.000	0.800	0.980
	τ_2	0.000	0.000	0.920	0.968	-0.002	-0.001	0.975	0.978
	τ_3	-0.027	-0.001	0.750	0.978	-0.017	-0.001	0.845	0.985
$n = 250$	τ_1	0.029	0.000	0.507	0.980	0.016	0.000	0.740	0.995
	τ_2	0.001	0.000	0.950	0.990	0.000	0.000	0.970	0.998
	τ_3	-0.027	-0.001	0.710	0.985	-0.015	0.000	0.838	0.998

Tableau 3.9: Bias and probability coverage in the location-shift and location-scale-shift scenario with the error term $\varepsilon \sim \chi^2_3$.

		$m = 5$				$m = 10$			
		Bias		$P_{0.95}$		Bias		$P_{0.95}$	
	τ	ERFE	QRFE	ERFE	QRFE	ERFE	QRFE	ERFE	QRFE
					Location-shift model				
$n = 50$	τ_1	0.001	-0.001	0.940	0.962	0.000	-0.001	0.978	0.985
	τ_2	0.001	0.000	0.970	0.972	0.000	0.000	0.985	0.985
	τ_3	0.001	0.001	0.980	0.970	0.000	0.001	0.990	0.985
$n = 100$	τ_1	0.000	-0.001	0.982	0.990	0.000	-0.002	0.995	0.990
	τ_2	0.000	0.000	0.990	0.988	0.000	0.000	0.995	0.995
	τ_3	0.000	0.002	0.992	0.975	0.000	0.001	0.995	0.992
$n = 250$	τ_1	0.000	-0.002	0.985	0.995	0.000	-0.002	1.000	0.995
	τ_2	0.000	0.000	0.995	1.000	0.000	0.000	1.000	0.998
	τ_3	0.000	0.001	0.998	0.982	0.000	0.001	1.000	0.998
					Location-scale-shift model				
$n = 50$	τ_1	0.029	0.000	0.628	0.952	0.017	0.001	0.740	0.978
	τ_2	0.001	0.001	0.890	0.948	0.000	0.000	0.945	0.975
	τ_3	-0.028	-0.002	0.802	0.955	-0.016	-0.001	0.848	0.972
$n = 100$	τ_1	0.028	-0.001	0.612	0.972	0.014	0.000	0.800	0.980
	τ_2	0.000	0.000	0.920	0.968	-0.002	-0.001	0.975	0.978
	τ_3	-0.027	-0.001	0.750	0.978	-0.017	-0.001	0.845	0.985
$n = 250$	τ_1	0.029	0.000	0.507	0.980	0.016	0.000	0.740	0.995
	τ_2	0.001	0.000	0.950	0.990	0.000	0.000	0.970	0.998
	τ_3	-0.027	-0.001	0.710	0.985	-0.015	0.000	0.838	0.998

Tableau 3.10: Parameters estimates (Est) with their 95% confidence intervals (CI) and p-values (P) obtained from the ERFE and QRFE methods at five percentiles, $\tau = (0.1, 0.25, 0.5, 0.75, 0.9)$.

	τ	0.1			0.25			0.5			0.75			0.9		
Var	Method	Est	CI	P	Est	CI	P	Est	CI	P	Est	CI	P	Est	CI	P
EXP	ERFE	0.11	(0.11, 0.11)	0.00	0.11	(0.11, 0.12)	0.00	0.11	(0.11, 0.12)	0.00	0.11	(0.11, 0.12)	0.00	0.11	(0.11, 0.12)	0.00
	QRFE	0.07	(0.06, 0.09)	0.00	0.07	(0.06, 0.08)	0.00	0.07	(0.06, 0.08)	0.00	0.08	(0.06, 0.09)	0.00	0.10	(0.08, 0.12)	0.00
EXP2	ERFE	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00
	QRFE	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00
IND	ERFE	0.03	(0.01, 0.06)	0.01	0.03	(0.00, 0.06)	0.08	0.02	(−0.01, 0.05)	0.23	0.01	(−0.02, 0.04)	0.53	0.01	(−0.03, 0.04)	0.70
	QRFE	0.07	(0.00, 0.15)	0.05	0.04	(−0.03, 0.11)	0.25	0.07	(−0.01, 0.16)	0.07	0.12	(0.03, 0.21)	0.01	0.14	(0.01, 0.26)	0.03
MS	ERFE	-0.05	(−0.07, −0.03)	0.00	-0.04	(−0.07, −0.01)	0.01	-0.03	(−0.06, 0.01)	0.09	-0.03	(−0.06, 0.01)	0.16	-0.03	(−0.07, 0.02)	0.25
	QRFE	0.19	(0.03, 0.36)	0.02	0.19	(0.06, 0.32)	0.00	0.32	(0.15, 0.48)	0.00	0.31	(0.15, 0.47)	0.00	0.30	(0.16, 0.44)	0.00
OCC	ERFE	-0.02	(−0.04, 0.00)	0.07	-0.02	(−0.04, 0.00)	0.10	-0.02	(−0.05, 0.01)	0.11	-0.02	(−0.05, 0.00)	0.08	-0.03	(−0.06, 0.00)	0.09
	QRFE	-0.02	(−0.11, 0.06)	0.58	0.04	(−0.04, 0.12)	0.36	0.10	(0.01, 0.20)	0.02	0.25	(0.15, 0.36)	0.00	0.47	(0.31, 0.63)	0.00
SMSA	ERFE	-0.05	(−0.08, −0.01)	0.00	-0.04	(−0.08, −0.01)	0.02	-0.04	(−0.08, −0.01)	0.02	-0.04	(−0.08, −0.01)	0.02	-0.04	(−0.09, 0.00)	0.06
	QRFE	0.15	(0.06, 0.24)	0.00	0.15	(0.08, 0.22)	0.00	0.16	(0.09, 0.23)	0.00	0.13	(0.05, 0.22)	0.00	0.22	(0.10, 0.34)	0.00
SOUTH	ERFE	-0.03	(−0.10, 0.04)	0.37	-0.02	(−0.12, 0.07)	0.61	0.00	(−0.12, 0.12)	0.98	0.03	(−0.14, 0.19)	0.75	0.03	(−0.18, 0.24)	0.76
	QRFE	-0.05	(−0.12, 0.02)	0.16	-0.05	(−0.13, 0.03)	0.18	0.04	(−0.04, 0.11)	0.35	0.11	(0.02, 0.20)	0.02	0.18	(0.04, 0.32)	0.01
UNION	ERFE	0.05	(0.03, 0.07)	0.00	0.04	(0.02, 0.07)	0.00	0.03	(0.00, 0.07)	0.05	0.02	(−0.01, 0.06)	0.22	0.01	(−0.03, 0.06)	0.48
	QRFE	0.26	(0.18, 0.34)	0.00	0.27	(0.20, 0.34)	0.00	0.23	(0.15, 0.30)	0.00	0.22	(0.13, 0.31)	0.00	0.25	(0.10, 0.40)	0.00
WKS	ERFE	0.00	(0.00, 0.00)	0.02	0.00	(0.00, 0.00)	0.03	0.00	(0.00, 0.00)	0.16	0.00	(0.00, 0.00)	0.56	0.00	(0.00, 0.00)	0.95
	QRFE	0.07	(0.06, 0.08)	0.00	0.07	(0.07, 0.08)	0.00	0.07	(0.06, 0.08)	0.00	0.07	(0.06, 0.07)	0.00	0.06	(0.05, 0.07)	0.00

3.7.2 Supplementary Material II : Proof of the Theorems

Proof of Theorem 3.3.1.

The proof of **Theorem 3.3.1** is adapted from a proof of Koenker (2004). First, we present a philosophical approach and, secondly, a thorough one. We adopt a purely heuristic approach and ignore the complications introduced by the infinite dimensional nature of the incidental parameter. In the completed version, we explicitly concentrate out the incidental parameter before showing the asymptotic property of the parameter estimator of interest. The Lemma 3.3.1.1 establishes the equivalence between the two approaches.

Part 1. Let $\mu_{ij\tau} = \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta}_{\tau} + \boldsymbol{z}_{ij}{}^{\mathsf{T}}\boldsymbol{\alpha}$ and consider the following objective function

$$R_{nm}(\boldsymbol{\delta}) = \sum_{i=1}^{n}\sum_{j=1}^{m}\rho_{\tau}\left\{y_{ij} - \mu_{ij\tau} - \boldsymbol{z}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_0/\sqrt{m} - \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_1/\sqrt{nm}\right\} - \rho_{\tau}\{y_{ij} - \mu_{ij\tau}\}. \quad (3.22)$$

The above objective function is a convex function of $\boldsymbol{\delta}$ that is minimized by

$$\widehat{\boldsymbol{\delta}} = \begin{pmatrix} \widehat{\boldsymbol{\delta}}_0 \\ \widehat{\boldsymbol{\delta}}_1 \end{pmatrix} = \begin{pmatrix} \sqrt{m}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \\ \sqrt{nm}(\widehat{\boldsymbol{\beta}}_{\tau} - \boldsymbol{\beta}_{\tau}) \end{pmatrix}. \quad (3.23)$$

Our goal is to approximate R_{nm} by a quadratic function with a unique minimizing value, and use results from Lid Hjort et Pollard (2011) to show that $\widehat{\boldsymbol{\delta}}$ has the same asymptotic distribution of that minimizing value. This quadratic approximation is mainly composed by the Taylor expansion of the expected value and by a linear approximation function.

Let $\widetilde{\boldsymbol{x}_{ij}} = (\boldsymbol{z}_{ij}{}^{\mathsf{T}}, \boldsymbol{x}_{ij}{}^{\mathsf{T}})^{\mathsf{T}}$, $\widetilde{\boldsymbol{\delta}} = (\boldsymbol{\delta}_0{}^{\mathsf{T}}/\sqrt{m}, \boldsymbol{\delta}_1{}^{\mathsf{T}}/\sqrt{nm})^{\mathsf{T}}$ and $\varepsilon_{ij\tau} = y_{ij} - \mu_{ij\tau}$. The function $\mathbb{E}(\rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}{}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau}))$ is convex, twice continuously differentiable and reaches its minimum at $\widetilde{\boldsymbol{\delta}} = \mathbf{0}$. It can be represented in the neighbourhood of $\widetilde{\boldsymbol{\delta}} = \mathbf{0}$ as

$$\begin{aligned}\mathbb{E}\left[\rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}{}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau})\right] = \widetilde{\boldsymbol{\delta}}^{\mathsf{T}}\widetilde{\boldsymbol{x}_{ij}}\,\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau})]\widetilde{\boldsymbol{x}_{ij}}{}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}} \\ - 2\widetilde{\boldsymbol{\delta}}^{\mathsf{T}}\widetilde{\boldsymbol{x}_{ij}}\,\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau}).\varepsilon_{ij\tau}] + o\left\|\widetilde{\boldsymbol{\delta}}\right\|^2,\end{aligned} \tag{3.24}$$

where $\psi_\tau(\lambda) = \tau - \mathbb{1}(\lambda < 0)$. Since

$$\underset{\boldsymbol{\delta} \in \mathbb{R}^{n+p}}{\operatorname{argmin}}\,\mathbb{E}\left[\rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}{}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau})\right] = \mathbf{0}$$

we have by the first order condition

$$\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau}).\varepsilon_{ij\tau}] = 0, \tag{3.25}$$

and equation (3.24) can be reduced to :

$$\mathbb{E}\left[\rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}{}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau})\right] = \widetilde{\boldsymbol{\delta}}^{\mathsf{T}}\widetilde{\boldsymbol{x}_{ij}}\,\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau})]\widetilde{\boldsymbol{x}_{ij}}{}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}} + o\left\|\widetilde{\boldsymbol{\delta}}\right\|^2. \tag{3.26}$$

The linear approximation function can be seen as a sort of Taylor expansion of $R_{nm}(\boldsymbol{\delta})$ around $\boldsymbol{\delta} = 0$, see (Pollard, 1991b). Define

$$D_{ij}(\varepsilon_{ij\tau}) = -2\psi_\tau(\varepsilon_{ij\tau}).\varepsilon_{ij\tau}. \tag{3.27}$$

Notice that by (3.25), $\mathbb{E}(D_{ij}(\varepsilon_{ij\tau})) = 0$. Define

$$r_{ij}(\widetilde{\boldsymbol{\delta}}) = \rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}^\mathsf{T}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau}) - \widetilde{\boldsymbol{\delta}}^\mathsf{T}\widetilde{\boldsymbol{x}_{ij}} D_{ij}(\varepsilon_{ij\tau}).$$

Then

$$\begin{aligned} R_{nm}(\widetilde{\boldsymbol{\delta}}) = &\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\mathbb{E}\left[\rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}^\mathsf{T}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau})\right]\right) \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{m}\widetilde{\boldsymbol{\delta}}^\mathsf{T}\widetilde{\boldsymbol{x}_{ij}} D_{ij}(\varepsilon_{ij\tau}) + \sum_{i=1}^{n}\sum_{j=1}^{m}\left(r_{ij}(\widetilde{\boldsymbol{\delta}}) - \mathbb{E}\left[r_{ij}(\widetilde{\boldsymbol{\delta}})\right]\right). \end{aligned}$$

Using Lemma 2.7.0.1, the objective function $R_{nm}(\widetilde{\boldsymbol{\delta}})$ reduce to

$$\begin{aligned} R_{nm}(\widetilde{\boldsymbol{\delta}}) &= \widetilde{\boldsymbol{\delta}}^\mathsf{T}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\widetilde{\boldsymbol{x}_{ij}}\,\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau})]\widetilde{\boldsymbol{x}_{ij}}^\mathsf{T}\right)\widetilde{\boldsymbol{\delta}} + \widetilde{\boldsymbol{\delta}}^\mathsf{T}\sum_{i=1}^{n}\sum_{j=1}^{m}\widetilde{\boldsymbol{x}_{ij}} D_{ij}(\varepsilon_{ij\tau}) + O\left\|\widetilde{\boldsymbol{\delta}}\right\|^2 + o\left\|\widetilde{\boldsymbol{\delta}}\right\|^2 \\ &= \widetilde{\boldsymbol{\delta}}^\mathsf{T}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\widetilde{\boldsymbol{x}_{ij}}\,\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau})]\widetilde{\boldsymbol{x}_{ij}}^\mathsf{T}\right)\widetilde{\boldsymbol{\delta}} + \widetilde{\boldsymbol{\delta}}^\mathsf{T}\sum_{i=1}^{n}\sum_{j=1}^{m}\widetilde{\boldsymbol{x}_{ij}} D_{ij}(\varepsilon_{ij\tau}) + o_p(1) \\ &\simeq \widetilde{\boldsymbol{\delta}}^\mathsf{T}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\widetilde{\boldsymbol{x}_{ij}}\,\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau})]\widetilde{\boldsymbol{x}_{ij}}^\mathsf{T}\right)\widetilde{\boldsymbol{\delta}} + \widetilde{\boldsymbol{\delta}}^\mathsf{T}\sum_{i=1}^{n}\sum_{j=1}^{m}\widetilde{\boldsymbol{x}_{ij}} D_{ij}(\varepsilon_{ij\tau}). \end{aligned} \tag{3.28}$$

By replacing $\widetilde{\boldsymbol{x}_{ij}} = (\boldsymbol{z}_{ij}^{\mathsf{T}}, \boldsymbol{x}_{ij}^{\mathsf{T}})^{\mathsf{T}}$ and $\widetilde{\boldsymbol{\delta}} = (\boldsymbol{\delta}_0^{\mathsf{T}}/\sqrt{m}, \boldsymbol{\delta}_1^{\mathsf{T}}/\sqrt{nm})^{\mathsf{T}}$ by their initial value, we have

$$
\begin{aligned}
&R_{nm}(\boldsymbol{\delta}) = \\
&-2\frac{1}{\sqrt{m}}\sum_{i=1}^{n}\sum_{j=1}^{m}(\boldsymbol{z}_{ij}^{\mathsf{T}}\boldsymbol{\delta}_0 + \boldsymbol{x}_{ij}^{\mathsf{T}}\boldsymbol{\delta}_1/\sqrt{n})\psi_\tau(y_{ij} - \mu_{ij\tau}).(y_{ij} - \mu_{ij\tau}) \\
&+\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb{E}[\psi_\tau(y_{ij} - \mu_{ij\tau})](\boldsymbol{z}_{ij}^{\mathsf{T}}\boldsymbol{\delta}_0 + \boldsymbol{x}_{ij}^{\mathsf{T}}\boldsymbol{\delta}_1/\sqrt{n})^2 \\
&= -2\frac{1}{\sqrt{m}}\left[\boldsymbol{\delta}_0^{\mathsf{T}}\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau + \boldsymbol{\delta}_1^{\mathsf{T}}/\sqrt{n}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau\right] \\
&+\frac{1}{m}\left[\boldsymbol{\delta}_0^{\mathsf{T}}\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{Z}\boldsymbol{\delta}_0 + 2\boldsymbol{\delta}_0^{\mathsf{T}}\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}\boldsymbol{\delta}_1/\sqrt{n} + \boldsymbol{\delta}_1^{\mathsf{T}}\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}\boldsymbol{\delta}_1/n\right] \\
&= R_{nm}^{(1)}(\boldsymbol{\delta}) + R_{nm}^{(2)}(\boldsymbol{\delta})
\end{aligned}
$$

The condition **A2** and **A3** imply a Lindberg condition and we have

$$R_{nm}^{(1)}(\boldsymbol{\delta}) = -2\frac{1}{\sqrt{m}}\left[\boldsymbol{\delta}_0^{\mathsf{T}}\boldsymbol{Z}^{\mathsf{T}} + \boldsymbol{\delta}_1^{\mathsf{T}}/\sqrt{n}\boldsymbol{X}^{\mathsf{T}}\right]\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau \xrightarrow{d} -2\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{B}.$$

While by condition **A2**

$$
\begin{aligned}
R_{nm}^{(2)}(\boldsymbol{\delta}) &= \frac{1}{m}\left[\boldsymbol{\delta}_0^{\mathsf{T}}\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{Z}\boldsymbol{\delta}_0 + 2\boldsymbol{\delta}_0^{\mathsf{T}}\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}\boldsymbol{\delta}_1/\sqrt{n} + \boldsymbol{\delta}_1^{\mathsf{T}}\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}\boldsymbol{\delta}_1/n\right] \\
&\to \boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{D}_1\boldsymbol{\delta}.
\end{aligned}
$$

Thus the limiting form of the objective function is

$$R_0(\boldsymbol{\delta}) = -2\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{B} + \boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{D}_1\boldsymbol{\delta}$$

where $\boldsymbol{B}$ is a zero mean Gaussian vector with covariance matrix $\boldsymbol{D}_0$. The objective function R_{nm} is convex and its limiting form R_0 has a unique minimum. Therefore,

by **Theorem**A.1, $\widehat{\boldsymbol{\delta}}$ converge to the argmin of R_0.

Part 2. In the precedent proof we overlook the complications related to the infinite dimensional nature of $\boldsymbol{\alpha}$. Koenker (2004), following (Bickel, 1975; Ruppert et Carroll, 1980), used the Bahadur-Kiefer representation of the incidental parameter in order to concentrate out its effect and express the objective function solely in terms of the finite dimensional parameter $\boldsymbol{\beta}$. With the smoothness of the asymmetric least square loss, we can use the first order condition to represent the incidental parameter as a function of the structural parameter.

Consider the following objective function

$$R_{nm}(\boldsymbol{\delta}) = -2\frac{1}{\sqrt{m}}\sum_{i=1}^{n}\sum_{j=1}^{m}(\boldsymbol{z}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_0 + \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_1/\sqrt{n})\psi_\tau(y_{ij}-\mu_{ij\tau}).(y_{ij}-\mu_{ij\tau})$$
$$+\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb{E}[\psi_\tau(y_{ij}-\mu_{ij\tau})](\boldsymbol{z}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_0 + \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_1/\sqrt{n})^2.$$

This function is twice derivable and the first order condition gives us the exact representation of the incidental parameter

$$\frac{\widehat{\delta}_{0i}}{\sqrt{m}} = \frac{1}{m\overline{\psi}_i}\sum_{k=1}^{m}\psi_\tau(y_{ik}-\mu_{ik\tau})(y_{ik}-\mu_{ik\tau}) - \frac{1}{m\overline{\psi}_i}\sum_{k=1}^{m}\mathbb{E}\left[\psi_\tau(y_{ik}-\mu_{ik\tau})\right]\boldsymbol{x}_{ik}{}^{\mathsf{T}}\frac{\boldsymbol{\delta}_1}{\sqrt{nm}},$$

where $\overline{\psi}_i = m^{-1}\sum_{k=1}^{m}\mathbb{E}\left[\psi_\tau\big(y_{ik}-\mu_{ik\tau}\big)\right]$. Substituting $\frac{\widehat{\delta}_{0i}}{\sqrt{m}}$ we have

$$\frac{\widehat{\boldsymbol{\delta}}_1}{\sqrt{mn}} = \left(\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{x}_{ij}\,\mathbb{E}[\psi_\tau(y_{ij}-\mu_{ij\tau})]\Big[\boldsymbol{x}_{ij}{}^{\mathsf{T}} - \frac{1}{m\overline{\psi}_i}\sum_{k=1}^{m}\mathbb{E}\left[\psi_\tau(y_{ik}-\mu_{ik\tau})\right]\boldsymbol{x}_{ik}{}^{\mathsf{T}}\Big]\right)^{-1}$$
$$\times\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{x}_{ij}\psi_\tau(y_{ij}-\mu_{ij\tau})(y_{ij}-\mu_{ij\tau})\right.$$
$$\left.-\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{x}_{ij}\,\mathbb{E}[\psi_\tau(y_{ij}-\mu_{ij\tau})]\Big[\frac{1}{m\overline{\psi}_i}\sum_{k=1}^{m}\psi_\tau(y_{ik}-\mu_{ik})(y_{ik}-\mu_{ik})\Big]\right).$$

Note that,

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{x}_{ij}\,\mathbb{E}[\psi_\tau(y_{ij}-\mu_{ij\tau})]\Big[\boldsymbol{x}_{ij}{}^{\mathsf{T}}-\frac{1}{m\overline{\psi}_i}\sum_{k=1}^{m}\mathbb{E}\left[\psi_\tau(y_{ik}-\mu_{ik})\right]\boldsymbol{x}_{ik}{}^{\mathsf{T}}\Big]\\
&=\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}-\boldsymbol{X}^{\mathsf{T}}\boldsymbol{P}_{\boldsymbol{Z}}{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}\\
&=\boldsymbol{X}^{\mathsf{T}}[\mathbb{I}-\boldsymbol{P}_{\boldsymbol{Z}}]\,\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}\\
&=\boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}\\
&=\boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X}\\
&=\boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{X}
\end{aligned}$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{x}_{ij}\psi_\tau(y_{ij}-\mu_{ij\tau})(y_{ij}-\mu_{ij\tau})-\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{x}_{ij}\,\mathbb{E}[\psi_\tau(y_{ij}-\mu_{ij\tau})]\times$$
$$\frac{1}{m\overline{\psi}_i}\sum_{k=1}^{m}\psi_\tau(y_{ik}-\mu_{ik\tau})(y_{ik}-\mu_{ik\tau})=\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau-\boldsymbol{X}^{\mathsf{T}}\boldsymbol{P}_{\boldsymbol{Z}}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau=\boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau.$$

Consequently,

$$\begin{aligned}
\frac{\widehat{\boldsymbol{\delta}}_1}{\sqrt{mn}}&=\Big(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{X}\Big)^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau\\
&=\Big(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{X}/m\Big)^{-1}m^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau.
\end{aligned}\tag{3.29}$$

Let $\boldsymbol{X}_{\boldsymbol{Z}}=\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{X}$ then we have $m^{-1}\boldsymbol{X}_{\boldsymbol{Z}}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau=\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{x}_{ij\boldsymbol{Z}}\psi_\tau(\varepsilon_{ij\tau})\varepsilon_{ij\tau}/m$. Let $T_{ni}=\sum_{j=1}^{m}m^{-1}\boldsymbol{\lambda}^{\mathsf{T}}\boldsymbol{x}_{ij\boldsymbol{Z}}\psi_\tau(\varepsilon_{ij\tau})\varepsilon_{ij\tau}$ and consider $n^{-1/2}\sum_{i=1}^{n}T_{ni}$, where $\boldsymbol{\lambda}$ is a $p\times 1$ unit vector, $\boldsymbol{\lambda}^{\mathsf{T}}\boldsymbol{\lambda}=1$. The summands T_{ni} are independent with $\mathbb{E}[T_{ni}]=0$ and $\mathrm{Var}\left[n^{-1/2}\sum_{i=1}^{n}T_{ni}\right]>\nu\prime>0$, by condition **A2**. By application of the Minkowski's

inequality, we have

$$\begin{aligned}
\mathbb{E}|T_{ni}|^{2+\nu} &= \mathbb{E}\left|\sum_{j=1}^{m}\sum_{k=1}^{p} m^{-1}\lambda_k x^k_{ij\boldsymbol{Z}}\psi_\tau(\varepsilon_{ij\tau})\varepsilon_{ij\tau}\right|^{2+\nu} \\
&\leq \left[\sum_{j=1}^{m}\sum_{k=1}^{p}\left(\mathbb{E}\left|m^{-1}\lambda_k x^k_{ij\boldsymbol{Z}}\psi_\tau(\varepsilon_{ij\tau})\varepsilon_{ij\tau}\right|^{2+\nu}\right)^{\frac{1}{2+\nu}}\right]^{2+\nu} \\
&\leq \left[\sum_{j=1}^{m}\sum_{k=1}^{p} m^{-1}|\lambda_k x^k_{ij\boldsymbol{Z}}|\left(\mathbb{E}\left|\psi_\tau(\varepsilon_{ij\tau})\varepsilon_{ij\tau}\right|^{2+\nu}\right)^{\frac{1}{2+\nu}}\right]^{2+\nu} \\
&\leq M\Delta p^{1+\nu},
\end{aligned}$$

where the last inequality follows by $\mathbb{E}|\psi_\tau(\varepsilon_{ij\tau})|^{4+\nu} < \Delta$ and $\mathbb{E}|\varepsilon_{ij\tau}|^{4+\nu} < \Delta$.

Then by the Liapounov CLT $\boldsymbol{X}^\top \boldsymbol{M}_{\boldsymbol{Z}}{}^\top \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau$ is Gaussian and by condition **A2**, $\widehat{\boldsymbol{\delta}}_1$ is zero mean Gaussian vector with covariance matrix $\mathrm{Var}(\widehat{\boldsymbol{\beta}}_\tau) = \widetilde{\boldsymbol{D}}_1^{-1}\widetilde{\boldsymbol{D}}_0\widetilde{\boldsymbol{D}}_1^{-1}$. □

Proof of Lemma 3.3.1.1.

we intend to show that this new covariance matrix $\widetilde{\boldsymbol{D}}_1^{-1}\widetilde{\boldsymbol{D}}_0\widetilde{\boldsymbol{D}}_1^{-1}$ is identical to the lower right diagonal block matrix $\boldsymbol{D}_1^{-1}\boldsymbol{D}_0\boldsymbol{D}_1^{-1}$.

We have $(\boldsymbol{D}_1^{-1}\boldsymbol{D}_0\boldsymbol{D}_1^{-1})_{22} = (\boldsymbol{D}_1^{-1})_{2\cdot}\boldsymbol{D}_0(\boldsymbol{D}_1^{-1})_{2\cdot}{}^\top$. Using the standard partitioned inverse formula for a general 2×2 partitioned matrix, we have

$$\begin{aligned}
mn(\boldsymbol{D}_1^{-1}\boldsymbol{D}_0\boldsymbol{D}_1^{-1})_{22} &= \begin{pmatrix} -\boldsymbol{F}\boldsymbol{E}^{-1} & \boldsymbol{E}^{-1}\end{pmatrix}\begin{pmatrix} \boldsymbol{Z}^\top\Sigma_\tau\boldsymbol{Z} & \boldsymbol{Z}^\top\Sigma_\tau\boldsymbol{X} \\ \boldsymbol{X}^\top\Sigma_\tau\boldsymbol{Z} & \boldsymbol{X}^\top\Sigma_\tau\boldsymbol{X}\end{pmatrix}\begin{pmatrix} -\boldsymbol{F}\boldsymbol{E}^{-1} \\ \boldsymbol{E}^{-1}\end{pmatrix} \\
&= \boldsymbol{E}^{-1}[\boldsymbol{F}^\top\boldsymbol{Z}^\top\Sigma_\tau\boldsymbol{Z}\boldsymbol{F} - \boldsymbol{X}^\top\Sigma_\tau\boldsymbol{Z}\boldsymbol{F} - \boldsymbol{F}^\top\boldsymbol{Z}^\top\Sigma_\tau\boldsymbol{X} + \boldsymbol{X}^\top\Sigma_\tau\boldsymbol{X}]\boldsymbol{E}^{-1}
\end{aligned}$$

where $\boldsymbol{E} = \boldsymbol{X}^\top\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X} - \boldsymbol{X}^\top\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{Z}(\boldsymbol{Z}^\top\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X} =$

$mn\widetilde{\boldsymbol{D}}_1$, and $\boldsymbol{ZF} = \boldsymbol{Z}(\boldsymbol{Z}^\mathsf{T}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{Z})^{-1}\boldsymbol{Z}^\mathsf{T}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{X} = \boldsymbol{P}_{\boldsymbol{Z}}\boldsymbol{X}$. The term in square brackets is

$$
\begin{aligned}
&\boldsymbol{F}^\mathsf{T}\boldsymbol{Z}^\mathsf{T}\boldsymbol{\Sigma}_\tau\boldsymbol{Z}\boldsymbol{F} - \boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}_\tau\boldsymbol{Z}\boldsymbol{F} - \boldsymbol{F}^\mathsf{T}\boldsymbol{Z}^\mathsf{T}\boldsymbol{\Sigma}_\tau\boldsymbol{X} + \boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}_\tau\boldsymbol{X} \\
&= \boldsymbol{X}^\mathsf{T}\boldsymbol{P}_{\boldsymbol{Z}}{}^\mathsf{T}\boldsymbol{\Sigma}_\tau\boldsymbol{P}_{\boldsymbol{Z}}\boldsymbol{X} - \boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}_\tau\boldsymbol{P}_{\boldsymbol{Z}}\boldsymbol{X} - \boldsymbol{X}^\mathsf{T}\boldsymbol{P}_{\boldsymbol{Z}}{}^\mathsf{T}\boldsymbol{\Sigma}_\tau\boldsymbol{X} + \boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}_\tau\boldsymbol{X} \\
&= \boldsymbol{X}^\mathsf{T}[\boldsymbol{P}_{\boldsymbol{Z}}{}^\mathsf{T}\boldsymbol{\Sigma}_\tau\boldsymbol{P}_{\boldsymbol{Z}} - \boldsymbol{\Sigma}_\tau\boldsymbol{P}_{\boldsymbol{Z}} - \boldsymbol{P}_{\boldsymbol{Z}}{}^\mathsf{T}\boldsymbol{\Sigma}_\tau + \boldsymbol{\Sigma}_\tau]\boldsymbol{X} \\
&= \boldsymbol{X}^\mathsf{T}[\mathbf{I} - \boldsymbol{P}_{\boldsymbol{Z}}{}^\mathsf{T}]\boldsymbol{\Sigma}_\tau[\mathbf{I} - \boldsymbol{P}_{\boldsymbol{Z}}]\boldsymbol{X} \\
&= \boldsymbol{X}^\mathsf{T}\boldsymbol{M}_{\boldsymbol{Z}}{}^\mathsf{T}\boldsymbol{\Sigma}_\tau\boldsymbol{M}_{\boldsymbol{Z}}\boldsymbol{X} \\
&= \widetilde{\boldsymbol{D}}_0.
\end{aligned}
$$

Finally, the result follows. □

Proof of Theorem 3.3.2.

In the following, we remove the asterix as an exponent to lighten the notation. Using the same approach to the proof of **Theorem** 3.3.1, the objective function $R_{mnq}(\boldsymbol{\delta})$ can be decomposed into two parts

$$R_{mnq}(\boldsymbol{\delta}) = R^{(1)}_{mnq}(\boldsymbol{\delta}) + R^{(2)}_{mnq}(\boldsymbol{\delta}).$$

$$
\begin{aligned}
R^{(1)}_{nmq}(\boldsymbol{\delta}) &= \frac{-2}{\sqrt{nm}} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{j=1}^{m} v_k \boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}} \boldsymbol{x}_{ij} \psi_{\tau_k}(\varepsilon_{ij\tau_k}) . \varepsilon_{ij\tau_k} \\
&= \frac{-2}{\sqrt{nm}} \sum_{k=1}^{q} v_k \boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}} \boldsymbol{X}^{\mathsf{T}} \boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{\tau_k}) \boldsymbol{\varepsilon}_{\tau_k} \\
&= \frac{-2}{\sqrt{nm}} \boldsymbol{\delta}^{\mathsf{T}} (\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}} \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{\boldsymbol{\tau}}) \boldsymbol{\varepsilon}_{\boldsymbol{\tau}} \\
&= \frac{-2}{\sqrt{nm}} \boldsymbol{\delta}^{\mathsf{T}} \begin{pmatrix} v_1 \boldsymbol{X}^{\mathsf{T}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & v_q \boldsymbol{X}^{\mathsf{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Psi}_{\tau_1}(\boldsymbol{\varepsilon}_{\tau_1}) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Psi}_{\tau_q}(\boldsymbol{\varepsilon}_{\tau_q}) \end{pmatrix} \begin{pmatrix} \boldsymbol{\varepsilon}_{\tau_1} \\ \vdots \\ \boldsymbol{\varepsilon}_{\tau_q} \end{pmatrix} \\
&= \frac{-2}{\sqrt{nm}} \boldsymbol{\delta}^{\mathsf{T}} \sum_{i=1}^{n} (\boldsymbol{V} \otimes \boldsymbol{X}_i)^{\mathsf{T}} \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}) \boldsymbol{\varepsilon}_{i\boldsymbol{\tau}} \\
&= \frac{-2}{\sqrt{nm}} \boldsymbol{\delta}^{\mathsf{T}} \sum_{i=1}^{n} \begin{bmatrix} v_1 \sum_{j=1}^{m} x^1_{ij} \psi_{\tau_1}(\varepsilon_{it\tau_1}) \varepsilon_{it\tau_1} \\ \vdots \\ v_1 \sum_{j=1}^{m} x^p_{ij} \psi_{\tau_1}(\varepsilon_{it\tau_1}) \varepsilon_{it\tau_1} \\ \vdots \\ v_q \sum_{j=1}^{m} x^1_{ij} \psi_{\tau_q}(\varepsilon_{it\tau_q}) \varepsilon_{it\tau_q} \\ \vdots \\ v_q \sum_{j=1}^{m} x^p_{ij} \psi_{\tau_q}(\varepsilon_{it\tau_q}) \varepsilon_{it\tau_q} \end{bmatrix} \\
&\xrightarrow{d} -2\boldsymbol{\delta}^{\mathsf{T}} \boldsymbol{B}.
\end{aligned}
$$

To show the asymptotic normality of $\boldsymbol{B}$, we apply the Cramér-Wold device and verify the Lyapunov's condition.

Let $T_{ni} = m^{-1} \boldsymbol{\lambda}^{\mathsf{T}} (\boldsymbol{V} \otimes \boldsymbol{X}_i)^{\mathsf{T}} \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}) \boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}$ and consider $n^{-1/2} \sum_{i=1}^{n} T_{ni}$, where $\boldsymbol{\lambda}$ is a $pq \times 1$ unit vector, $\boldsymbol{\lambda}^{\mathsf{T}} \boldsymbol{\lambda} = 1$. The summands T_{ni} are independent with $\mathbb{E}[T_{ni}] = 0$

and $\mathrm{Var}\left[n^{-1/2}\sum_{i=1}^{n} T_{ni}\right] > \nu\prime > 0$, by condition **A2**. By the Minkowski's inequality, we have

$$\begin{aligned}
\mathbb{E}|T_{ni}|^{2+\nu} &= \mathbb{E}\left|\sum_{k=1}^{q}\sum_{l=1}^{p} v_k\lambda_{kl}\sum_{j=1}^{m} m^{-1}x_{ij}^{l}\psi_{\tau_k}(\varepsilon_{ij\tau_k})\varepsilon_{ij\tau_k}\right|^{2+\nu} \\
&= \mathbb{E}\left|\sum_{k=1}^{q}\sum_{l=1}^{p}\sum_{j=1}^{m} m^{-1}v_k\lambda_{kl}x_{ij}^{l}\psi_{\tau_k}(\varepsilon_{ij\tau_k})\varepsilon_{ij\tau_k}\right|^{2+\nu} \\
&\leq \left[\sum_{k=1}^{q}\sum_{l=1}^{p}\sum_{j=1}^{m}\left(\mathbb{E}\left|m^{-1}v_k\lambda_{kl}x_{ij}^{l}\psi_{\tau_k}(\varepsilon_{ij\tau_k})\varepsilon_{ij\tau_k}\right|^{2+\nu}\right)^{\frac{1}{2+\nu}}\right]^{2+\nu}.
\end{aligned}$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\mathbb{E}\left|m^{-1}v_k\lambda_{kl}x_{ij}^{l}\psi_{\tau_k}(\varepsilon_{ij\tau_k})\varepsilon_{ij\tau_k}\right|^{2+\nu} &= |m^{-1}v_k\lambda_{kl}x_{ij}^{l}|^{2+\nu}\,\mathbb{E}\left|\psi_{\tau_k}(\varepsilon_{ij\tau_k})\varepsilon_{ij\tau_k}\right|^{2+\nu} \\
&\leq (m^{-1}M|\lambda_{kl}|)^{2+\nu}\left[\mathbb{E}|\psi_{\tau_k}(\varepsilon_{ij\tau_k})|^{4+\nu}\right]^{1/2}\left[\mathbb{E}|\varepsilon_{ij\tau_k}|^{4+\nu}\right]^{1/2} \\
&\leq (m^{-1}M|\lambda_{kl}|)^{2+\nu}\Delta,
\end{aligned}$$

where the last inequality follows by $\mathbb{E}|\psi_{\tau_k}(\varepsilon_{ij\tau_k})|^{4+\nu} < \Delta$ and $\mathbb{E}|\varepsilon_{ij\tau_k}|^{4+\nu} < \Delta$. Therefore,

$$\begin{aligned}
\mathbb{E}|T_{ni}|^{2+\nu} &\leq \left[\sum_{k=1}^{q}\sum_{l=1}^{p}\sum_{j=1}^{m} m^{-1}M|\lambda_{kl}|\Delta^{\frac{1}{2+\nu}}\right]^{2+\nu} \\
&\leq \Delta M^{2+\nu}|\boldsymbol{\lambda}^{\mathsf{T}}\mathbb{1}_{pq}|^{2+\nu} \\
&\leq \Delta(pq)^{1+\nu}.
\end{aligned}$$

Then by the Liapounov CLT $\boldsymbol{B}$ is a zero mean Gaussian vector with covariance matrix $\boldsymbol{D}_0(\boldsymbol{\tau})$.

$$
\begin{aligned}
R^{(2)}_{nmq}(\boldsymbol{\delta}) &= \frac{1}{nm}\sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}}\boldsymbol{x}_{ij}\,\mathbb{E}[\psi_{\tau_k}(\varepsilon_{ij\tau_k})]\boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_{\tau_k} \\
&= \frac{1}{nm}\sum_{k=1}^{q} v_k \boldsymbol{\delta}_{\tau_k}{}^{\mathsf{T}}\boldsymbol{X}^{\mathsf{T}}\,\mathbb{E}[\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{\tau_k})]\boldsymbol{X}\boldsymbol{\delta}_{\tau_k} \\
&\to \boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{D}_1(\boldsymbol{\tau})\boldsymbol{\delta}.
\end{aligned}
$$

Thus the limiting form of the objective function is

$$R_{0q}(\boldsymbol{\delta}) = -2\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{B} + \boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{D}_1(\boldsymbol{\tau})\boldsymbol{\delta}$$

where $\boldsymbol{B}$ is a zero mean Gaussian vector with covariance matrix $\boldsymbol{D}_0(\boldsymbol{\tau})$. Application of Theorem 2.2 of Lid Hjort et Pollard (2011) gives the result of Theorem 3.3.2. □

Proof of Theorem 3.3.3.

We have

$$
\begin{aligned}
\widehat{\boldsymbol{D}}_1(\boldsymbol{\tau}) &= \frac{1}{nm}(\mathbf{I}_q \otimes \widehat{\boldsymbol{X}^*})^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}})(\boldsymbol{V}\otimes\widehat{\boldsymbol{X}^*}) \\
&= \frac{1}{nm}\sum_{i=1}^{n}\operatorname{diag}\left(v_1\widehat{\boldsymbol{X}^*_i}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_1}(\widehat{\boldsymbol{\varepsilon}^*_{i\tau_1}})\widehat{\boldsymbol{X}^*_i},\ldots,v_q\widehat{\boldsymbol{X}^*_i}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_q}(\widehat{\boldsymbol{\varepsilon}^*_{i\tau_q}})\widehat{\boldsymbol{X}^*_i}\right).
\end{aligned}
$$

The convergence of $\widehat{\boldsymbol{D}}_1(\boldsymbol{\tau})$ is obtained by showing the convergence of the general term $\frac{1}{nm}.\sum_{i=1}^{n} v_k\widehat{\boldsymbol{X}^*_i}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}^*_{i\tau_k}})\widehat{\boldsymbol{X}^*_i}$. This general term breaks down as follows :

$$
\begin{aligned}
\sum_{i=1}^{n}\widehat{\boldsymbol{X}^*_i}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^*_{i\tau}})\widehat{\boldsymbol{X}^*_i} &= \widehat{\boldsymbol{X}^*}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^*_{\tau}})\widehat{\boldsymbol{X}^*} = \boldsymbol{X}^{\mathsf{T}}\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^*_{\tau}})\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)\boldsymbol{X} \\
&= \boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^*_{\tau}})\widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)\boldsymbol{X} = \boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^*_{\tau}})[\mathbf{I}_{nm} - \widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau)]\boldsymbol{X} \\
&= \boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^*_{\tau}})\boldsymbol{X} - \boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau}(\widehat{\boldsymbol{\varepsilon}^*_{\tau}})\widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}.
\end{aligned}
\tag{3.30}
$$

We will show the convergence of each of the terms separately. First, consider the following expression : $|\psi_{\tau}(\widehat{\varepsilon}^*_{ij\tau}) - \psi_{\tau}(\varepsilon^*_{ij\tau})|$. This expression is 0 except when $\boldsymbol{x}^*_{ij}{}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_{\tau} \leq$

$y_{ij}^* \leq \boldsymbol{x}_{ij}^{*\mathsf{T}}\boldsymbol{\beta}_\tau$ or $\boldsymbol{x}_{ij}^{*\mathsf{T}}\boldsymbol{\beta}_\tau \leq y_{ij}^* \leq \boldsymbol{x}_{ij}^{*\mathsf{T}}\widehat{\boldsymbol{\beta}}_\tau$. It can be bounded as follows :

$$\begin{aligned}
|\psi_\tau(\widehat{\varepsilon}_{ij\tau}^*) - \psi_\tau(\varepsilon_{ij\tau}^*)| &\leq |1-2\tau| \mathbb{1}\big(|\varepsilon_{ij\tau}^*| \leq |\boldsymbol{x}_{ij}^{*\mathsf{T}}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)|\big) \\
&\leq |1-2\tau| \mathbb{1}\big(|\varepsilon_{ij\tau}^*| \leq \|\boldsymbol{x}_{ij}^*\| \left\|\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau\right\|\big) \\
&\leq |1-2\tau| \mathbb{1}\big(|\varepsilon_{ij\tau}^*| \leq pM \left\|\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau\right\|\big).
\end{aligned} \tag{3.31}$$

As plim $\widehat{\boldsymbol{\beta}}_\tau = \boldsymbol{\beta}_\tau$, we have, by equation (3.31) and Markov's inequality :

$$(nm)^{-1}\boldsymbol{X}^\mathsf{T}[\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_\tau^*) - \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau^*)]\boldsymbol{X} = (nm)^{-1}\sum_{i=1}^n \boldsymbol{X}_i{}^\mathsf{T}[\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}_{i\tau}^*}) - \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}^*)]\boldsymbol{X}_i \xrightarrow{p} \mathbf{0}.$$

In other words :

$$(nm)^{-1}\boldsymbol{X}^\mathsf{T}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}}_\tau^*)\boldsymbol{X} = (nm)^{-1}\boldsymbol{X}^\mathsf{T}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau^*)\boldsymbol{X} + o_p(1). \tag{3.32}$$

This result (3.32) is important and will be used frequently below. Another useful result is the convergence of the function $\widehat{w}_{i\tau}^* = (\sum_{j=1}^m \psi_\tau(\widehat{\varepsilon}_{ij\tau}^*))^{-1}$ which appears in the expression of $\widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau)$. This function is bounded by :

$$m\min(\tau, 1-\tau) \leq \widehat{w}_{i\tau}^{*-1} \leq m\max(\tau, 1-\tau).$$

We have

$$\begin{aligned}
\Big|\frac{1}{nm}\sum_{i=1}^n \widehat{w}_{i\tau}^* - \frac{1}{nm}\sum_{i=1}^n \mathbb{E}[w_{i\tau}^*]\Big| &\leq \Big|\frac{1}{nm}\sum_{i=1}^n w_{i\tau}^* - \frac{1}{nm}\sum_{i=1}^n \mathbb{E}[w_{i\tau}^*]\Big| \\
&\quad + \Big|\frac{1}{nm}\sum_{i=1}^n \widehat{w}_{i\tau}^* - \frac{1}{nm}\sum_{i=1}^n w_{i\tau}^*\Big|.
\end{aligned} \tag{3.33}$$

The first term converges by Markov's Law of Large Numbers (LLN). The second term is bounded by :

$$\Big|\frac{1}{nm}\sum_{i=1}^n \widehat{w}_{i\tau}^* - \frac{1}{nm}\sum_{i=1}^n w_{i\tau}^*\Big| \leq \frac{1}{nm^3}\sum_{i=1}^n\sum_{j=1}^m |\psi_\tau(\widehat{\varepsilon}_{ij\tau}^*) - \psi_\tau(\varepsilon_{ij\tau}^*)|.$$

Thus, convergence is achieved by the application of (3.31).

Now, let's show the convergence of the first term on the last line of equation (3.30). We have :

$$\begin{aligned}
&\left|\frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}^*_{i\tau}})\boldsymbol{X}_i-\frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}^*_{i\tau})]\boldsymbol{X}_i\right| \\
&\quad\leq\left|\frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}[\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}^*_{i\tau}})-\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}^*_{i\tau})]\boldsymbol{X}_i\right| \\
&\quad+\left|\frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}^*_{i\tau})\boldsymbol{X}_i-\frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i{}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}^*_{i\tau})]\boldsymbol{X}_i\right|.
\end{aligned}\tag{3.34}$$

The result (3.32) gives the convergence of the first term on the right of the inequality and the application of Markov's Law of large-number (LLN) gives the convergence of the second term.

The second term on the last line of equation (3.30) is expressed by :

$$\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\widehat{\boldsymbol{\varepsilon}^*_\tau})\widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}=\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{m}\widehat{w}^*_{i\tau}\psi_\tau(\widehat{\varepsilon}^*_{ij\tau})\psi_\tau(\widehat{\varepsilon}^*_{ik\tau})\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}.\tag{3.35}$$

We will use the following relation to show its convergence

$$\begin{aligned}
\widehat{w}^*_{i\tau}\psi_\tau(\widehat{\varepsilon}^*_{ij\tau})\psi_\tau(\widehat{\varepsilon}^*_{ik\tau})-w^*_{i\tau}\psi_\tau(\varepsilon^*_{ij\tau})\psi_\tau(\varepsilon^*_{ik\tau})&=(\widehat{w}^*_{i\tau}-w^*_{i\tau})\psi_\tau(\widehat{\varepsilon}^*_{ij\tau})\psi_\tau(\widehat{\varepsilon}^*_{ik\tau}) \\
&+w^*_{i\tau}\big\{\psi_\tau(\widehat{\varepsilon}^*_{ij\tau})[\psi_\tau(\widehat{\varepsilon}^*_{ik\tau})-\psi_\tau(\varepsilon^*_{ik\tau})]+\psi_\tau(\varepsilon^*_{ik\tau})[\psi_\tau(\widehat{\varepsilon}^*_{ij\tau})-\psi_\tau(\varepsilon^*_{ij\tau})]\big\}.
\end{aligned}\tag{3.36}$$

To show the convergence of the expression (3.35), consider :

$$
\begin{aligned}
&\Big|\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}\widehat{w}^{*}_{i\tau}\psi_{\tau}(\widehat{\varepsilon}^{*}_{ij\tau})\psi_{\tau}(\widehat{\varepsilon}^{*}_{ik\tau})-\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}\,\mathbb{E}[w^{*}_{i\tau}\psi_{\tau}(\varepsilon^{*}_{ij\tau})\psi_{\tau}(\varepsilon^{*}_{ik\tau})]\Big| \\
&\quad\leq\Big|\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}[\widehat{w}^{*}_{i\tau}\psi_{\tau}(\widehat{\varepsilon}^{*}_{ij\tau})\psi_{\tau}(\widehat{\varepsilon}^{*}_{ik\tau})-w^{*}_{i\tau}\psi_{\tau}(\varepsilon^{*}_{ij\tau})\psi_{\tau}(\varepsilon^{*}_{ik\tau})]\Big| \\
&\quad+\Big|\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}w^{*}_{i\tau}\psi_{\tau}(\varepsilon^{*}_{ij\tau})\psi_{\tau}(\varepsilon^{*}_{ik\tau})-\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}\,\mathbb{E}[w^{*}_{i\tau}\psi_{\tau}(\varepsilon^{*}_{ij\tau})\psi_{\tau}(\varepsilon^{*}_{ik\tau})]\Big| \\
&\quad\leq\Big|\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}(\widehat{w}^{*}_{i\tau}-w^{*}_{i\tau})\psi_{\tau}(\widehat{\varepsilon}^{*}_{ij\tau})\psi_{\tau}(\widehat{\varepsilon}^{*}_{ik\tau})\Big| \\
&\quad+\Big|\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}w^{*}_{i\tau}\{\psi_{\tau}(\widehat{\varepsilon}^{*}_{ij\tau})[\psi_{\tau}(\widehat{\varepsilon}^{*}_{ik\tau})-\psi_{\tau}(\varepsilon^{*}_{ik\tau})]+\psi_{\tau}(\varepsilon^{*}_{ik\tau})[\psi_{\tau}(\widehat{\varepsilon}^{*}_{ij\tau})-\psi_{\tau}(\varepsilon^{*}_{ij\tau})]\}\Big| \\
&\quad+\Big|\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}w^{*}_{i\tau}\psi_{\tau}(\varepsilon^{*}_{ij\tau})\psi_{\tau}(\varepsilon^{*}_{ik\tau})-\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}\,\mathbb{E}[w^{*}_{i\tau}\psi_{\tau}(\varepsilon^{*}_{ij\tau})\psi_{\tau}(\varepsilon^{*}_{ik\tau})]\Big| \\
&\quad\leq\Big|\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}(\widehat{w}^{*}_{i\tau}-w^{*}_{i\tau})\Big|+\Big|\frac{1}{nm^{2}}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}[\psi_{\tau}(\widehat{\varepsilon}^{*}_{ik\tau})-\psi_{\tau}(\varepsilon^{*}_{ik\tau})]\Big| \\
&\quad+\Big|\frac{1}{nm^{2}}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}[\psi_{\tau}(\widehat{\varepsilon}^{*}_{ij\tau})-\psi_{\tau}(\varepsilon^{*}_{ij\tau})]\Big| \\
&\quad+\Big|\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}w^{*}_{i\tau}\psi_{\tau}(\varepsilon^{*}_{ij\tau})\psi_{\tau}(\varepsilon^{*}_{ik\tau})-\frac{1}{nm}\sum_{i=1}^{n}\sum_{jk}^{m}\boldsymbol{x}_{ij}\boldsymbol{x}_{ik}{}^{\mathsf{T}}\,\mathbb{E}[w^{*}_{i\tau}\psi_{\tau}(\varepsilon^{*}_{ij\tau})\psi_{\tau}(\varepsilon^{*}_{ik\tau})]\Big|.
\end{aligned}
\tag{3.37}
$$

Finally, the convergence of the expression (3.35) results from the application of the relations (3.32) and (3.33) and from the application of Markov's Law of Large Numbers.

The second term of the variance-covariance matrix whose convergence we must show

is :

$$\widehat{\boldsymbol{D}}_0(\boldsymbol{\tau}) = \frac{1}{nm}(\boldsymbol{V} \otimes \widehat{\boldsymbol{X}^*})^{\mathsf{T}} \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}}) \widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}} \widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}}^{\mathsf{T}} \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\widehat{\boldsymbol{\varepsilon}^*_{\boldsymbol{\tau}}})(\boldsymbol{V} \otimes \widehat{\boldsymbol{X}^*}),$$

$$= \frac{1}{nm} \sum_{i=1}^{n} \begin{pmatrix} v_1^2 \widehat{\boldsymbol{X}}_i^{*\mathsf{T}} \widehat{\boldsymbol{\Sigma}}^*_{i\tau_1\tau_1} \widehat{\boldsymbol{X}}_i^* & \cdots & \cdots & v_1 v_q \widehat{\boldsymbol{X}}_i^{*\mathsf{T}} \widehat{\boldsymbol{\Sigma}}^*_{i\tau_1\tau_q} \widehat{\boldsymbol{X}}_i^* \\ \vdots & \vdots & \ddots & \vdots \\ v_1 v_q \widehat{\boldsymbol{X}}_i^{*\mathsf{T}} \widehat{\boldsymbol{\Sigma}}^*_{i\tau_q\tau_1} \widehat{\boldsymbol{X}}_i^* & \cdots & \cdots & v_q^2 \widehat{\boldsymbol{X}}_i^{*\mathsf{T}} \widehat{\boldsymbol{\Sigma}}^*_{i\tau_q\tau_q} \widehat{\boldsymbol{X}}_i^* \end{pmatrix}$$

where

$$\widehat{\boldsymbol{\Sigma}}^*_{i\tau_k\tau_j} = \boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{i\tau_k}) \widehat{\boldsymbol{\varepsilon}}^*_{i\tau_k} \widehat{\boldsymbol{\varepsilon}}^*_{i\tau_j}{}^{\mathsf{T}} \boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{i\tau_j}).$$

Again, showing the convergence of the general term suffices :

$$\frac{1}{nm} \sum_{i=1}^{n} \widehat{\boldsymbol{X}}_i^{*\mathsf{T}} \widehat{\boldsymbol{\Sigma}}^*_{i\tau_k\tau_j} \widehat{\boldsymbol{X}}_i^* = \frac{1}{nm} \widehat{\boldsymbol{X}^*}^{\mathsf{T}} \boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k}) \widehat{\boldsymbol{\varepsilon}}^*_{\tau_k} \widehat{\boldsymbol{\varepsilon}}^*_{\tau_j}{}^{\mathsf{T}} \boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j}) \widehat{\boldsymbol{X}^*} \tag{3.38}$$

Note that,

$$\widehat{\boldsymbol{\varepsilon}}^*_{\tau} = \boldsymbol{\varepsilon}^*_{\tau} - \boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau} - \boldsymbol{\beta}_{\tau}) + \Delta \boldsymbol{M}_{\boldsymbol{Z}}(\tau)[\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_{\tau}],$$

where $\Delta \boldsymbol{M}_{\boldsymbol{Z}}(\tau) = \widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau) - \boldsymbol{M}_{\boldsymbol{Z}}(\tau) = \boldsymbol{P}_{\boldsymbol{Z}}(\tau) - \widehat{\boldsymbol{P}}_{\boldsymbol{Z}}(\tau) = \Delta \boldsymbol{P}_{\boldsymbol{Z}}(\tau)$. We have,

$$\begin{aligned} \widehat{\boldsymbol{\varepsilon}}^*_{\tau_k} \widehat{\boldsymbol{\varepsilon}}^*_{\tau_j}{}^{\mathsf{T}} &= \boldsymbol{\varepsilon}^*_{\tau_k} \boldsymbol{\varepsilon}^*_{\tau_j}{}^{\mathsf{T}} - \boldsymbol{\varepsilon}^*_{\tau_k}(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}} \boldsymbol{X}^{*\mathsf{T}} + \boldsymbol{\varepsilon}^*_{\tau_k}[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}} \Delta \boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}} \\ &\quad - \boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k}) \boldsymbol{\varepsilon}^*_{\tau_j}{}^{\mathsf{T}} + \boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}} \boldsymbol{X}^{*\mathsf{T}} \\ &\quad - \boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}} \Delta \boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}} + \Delta \boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}] \boldsymbol{\varepsilon}^*_{\tau_j}{}^{\mathsf{T}} \\ &\quad - \Delta \boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}](\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}} \boldsymbol{X}^{*\mathsf{T}} \\ &\quad + \Delta \boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}][\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}} \Delta \boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}} \end{aligned}$$

Then, replacing $\widehat{\boldsymbol{X}^*} = \widehat{\boldsymbol{M}}_{\boldsymbol{Z}}(\tau)\boldsymbol{X} = \boldsymbol{X}^* + \Delta \boldsymbol{M}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}$, we obtain :

$$
\begin{aligned}
&\widehat{\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k}\widehat{\boldsymbol{\varepsilon}}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^*} = \\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* - \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&- \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&- \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}]\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&- \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}](\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}][\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \\
&- \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \\
&- \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \\
&- \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}]\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \\
&- \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}](\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}][\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \\
&+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \\
&+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^*
\end{aligned}
$$

(continued on next page)

(continued from previous page)

$$
\begin{aligned}
&- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^*\\
&+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}]\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^*\\
&- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}](\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^*\\
&+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}][\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^*\\
&+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X}\\
&- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}(\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X}\\
&+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X}\\
&- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X}\\
&+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})(\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X}\\
&- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X}\\
&+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}]\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X}\\
&- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}](\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X}\\
&+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}][\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X}.
\end{aligned}
$$

Now let's break down the following expression :

$$
\begin{aligned}
\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* &= \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\boldsymbol{\varepsilon}^*_{\tau_j})\boldsymbol{X}^*\\
&+ \boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\Big[\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})-\boldsymbol{\Psi}_{\tau_j}(\boldsymbol{\varepsilon}^*_{\tau_j})\Big]\boldsymbol{X}^*\\
&+ \boldsymbol{X}^{*\mathsf{T}}\Big[\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})-\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{\tau_k})\Big]\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\boldsymbol{\varepsilon}^*_{\tau_j})\boldsymbol{X}^*\\
&+ \boldsymbol{X}^{*\mathsf{T}}\Big[\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})-\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{\tau_k})\Big]\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\Big[\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})-\boldsymbol{\Psi}_{\tau_j}(\boldsymbol{\varepsilon}^*_{\tau_j})\Big]\boldsymbol{X}^*.
\end{aligned}
$$

The final expression of (3.38) is :

$$\widehat{\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k}\widehat{\boldsymbol{\varepsilon}}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^*} =$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\boldsymbol{\varepsilon}^*_{\tau_j})\boldsymbol{X}^* \tag{e1}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\Big[\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})-\boldsymbol{\Psi}_{\tau_j}(\boldsymbol{\varepsilon}^*_{\tau_j})\Big]\boldsymbol{X}^* \tag{e2}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\Big[\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})-\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{\tau_k})\Big]\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\boldsymbol{\varepsilon}^*_{\tau_j})\boldsymbol{X}^* \tag{e3}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\Big[\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})-\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{\tau_k})\Big]\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\Big[\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})-\boldsymbol{\Psi}_{\tau_j}(\boldsymbol{\varepsilon}^*_{\tau_j})\Big]\boldsymbol{X}^* \tag{e4}$$

$$-\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}(\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e5}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e6}$$

$$-\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e7}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})(\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e8}$$

$$-\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e9}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}]\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e10}$$

$$-\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}](\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e11}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}][\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e12}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e13}$$

$$-\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}(\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e14}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e15}$$

$$-\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e16}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})(\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e17}$$

$$-\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e18}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}]\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e19}$$

$$-\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}](\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e20}$$

$$+\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}][\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e21}$$

(continued on next page)

(continued from previous page)

$$+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e22}$$

$$- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e23}$$

$$+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e24}$$

$$- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e25}$$

$$+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e26}$$

$$- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e27}$$

$$+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}]\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e28}$$

$$- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}](\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e29}$$

$$+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}][\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\boldsymbol{X}^* \tag{e30}$$

$$+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e31}$$

$$- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}(\widehat{\boldsymbol{\beta}}_{\tau_j} - \boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e32}$$

$$+ \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{\varepsilon}^*_{\tau_k}[\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e33}$$

$$- \boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_{\tau_k})\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e34}$$

(continued on next page)

(continued from previous page)

$$+\boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})(\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e35}$$

$$-\boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\boldsymbol{X}^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e36}$$

$$+\boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}]\boldsymbol{\varepsilon}^{*\mathsf{T}}_{\tau_j}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e37}$$

$$-\boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}](\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X} \tag{e38}$$

$$+\boldsymbol{X}^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_k)[\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_k}][\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\tau_j}]^{\mathsf{T}}\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_j})\Delta\boldsymbol{P}_{\boldsymbol{Z}}(\tau_j)\boldsymbol{X}. \tag{e39}$$

The demonstration is based on showing the convergence of each of these terms (e2)-(e39). We have three types of expression : those that are function of $\Delta\boldsymbol{P}_{\boldsymbol{Z}}$, those that are function of $\left[\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})-\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{\tau_k})\right]$ and those that are function of $(\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})$. Those that are a function of $\Delta\boldsymbol{P}_{\boldsymbol{Z}}$ are shown by following the approach of (3.33) and those that are function of $\left[\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{\tau_k})-\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{\tau_k})\right]$ according to the approach (3.32). The convergence technique of those that are a function of $(\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})$ is identical to the convergence procedure used for (e7) and (e8). Thus, in the following, we will show the convergence of (e7) and (e8).

Using the relation, $\mathrm{Vec}(ABC) = (C^{\mathsf{T}}\otimes A)\,\mathrm{Vec}(B)$, equation (e7) is transformed

as follows :

$$\mathrm{Vec}\Big(\frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k}^{*})\boldsymbol{X}_i^{*}(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})\boldsymbol{\varepsilon}_{i\tau_j}^{*\,\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_j}^{*})\boldsymbol{X}_i^{*}\Big)$$
$$=\frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k}^{*})\boldsymbol{\varepsilon}_{i\tau_j}^{*}\otimes\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_j})\boldsymbol{X}_i^{*}\,\mathrm{Vec}(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k}).$$

Applying **Lemma A.1**, we have

$$\mathbb{E}\left\|\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k}^{*})\boldsymbol{\varepsilon}_{i\tau_j}^{*}\otimes\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_j})\boldsymbol{X}_i^{*}\right\|^{1+\nu}$$
$$\leq\Big(\mathbb{E}\left\|\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k}^{*})\boldsymbol{\varepsilon}_{i\tau_j}^{*}\right\|^{2+2\nu}\mathbb{E}\left\|\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_j})\boldsymbol{X}_i^{*}\right\|^{2+2\nu}\Big)^{1/2}.$$

Repeated application of Minkowski and Holder inequalities shows that

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k}^{*})\boldsymbol{\varepsilon}_{i\tau_j}^{*}\right\|^{2+2\nu}&=\mathbb{E}\Big[\sum_{k=1}^{p}\Big|\sum_{j=1}^{m}x_{ij}^{*k}\psi_{\tau_k}(\widehat{\varepsilon}_{ij\tau_k}^{*})\varepsilon_{ij\tau_j}^{*}\Big|^{2}\Big]^{1+\nu}\\&\leq\Big[\sum_{k=1}^{p}\Big(\mathbb{E}\Big|\sum_{j=1}^{m}x_{ij}^{*k}\psi_{\tau_k}(\widehat{\varepsilon}_{ij\tau_k}^{*})\varepsilon_{ij\tau_j}^{*}\Big|^{2+2\nu}\Big)^{\frac{1}{1+\nu}}\Big]^{1+\nu}\\&\leq\Big\{\sum_{k=1}^{p}\Big[\sum_{j=1}^{m}\Big(\mathbb{E}|x_{ij}^{*k}\psi_{\tau_k}(\widehat{\varepsilon}_{ij\tau_k}^{*})\varepsilon_{ij\tau_j}^{*}|^{2+2\nu}\Big)^{\frac{1}{2+2\nu}}\Big]^{2+2\nu}\Big\}^{1+\nu}\\&\leq(p\Delta)^{1+\nu}(Mm)^{2+\nu}.\end{aligned}$$

The last inequality is obtained by applying the assumptions $\mathbb{E}|\psi_\tau(\widehat{\varepsilon}_{ij\tau})|^{4+\nu}<\Delta$ and $\mathbb{E}|\varepsilon_{ij\tau}|^{4+\nu}<\Delta$. Similarly

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_j})\boldsymbol{X}_i^{*}\right\|^{2+2\nu}&\leq\mathbb{E}\left\|\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}^{1/2}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_j})\right\|^{4+4\nu}\\&\leq\Big[\sum_{k=1}^{p}\sum_{j=1}^{m}\Big(\mathbb{E}|(x_{ij}^{*k}\psi_{\tau_j}^{1/2}(\widehat{\varepsilon}_{ij\tau_j}^{*}))^{2}|^{2+2\nu}\Big)^{\frac{1}{2+2\nu}}\Big]^{2+2\nu}\\&\leq(pm)^{2+\nu}\Delta.\end{aligned}$$

Then, by the Markov LLN it follows that

$$\mathrm{Vec}\Big(\frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_k}^{*})\boldsymbol{X}_i^{*}(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})\boldsymbol{\varepsilon}_{i\tau_j}^{*\,\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}_{i\tau_j}^{*})\boldsymbol{X}_i^{*}\Big)=\frac{1}{\sqrt{nm}}O_p(1)O_p(1).$$

Considering that $\text{Vec}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau) = O_p((nm)^{-1/2})$ and that $\mathbb{E}\left\|\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{i\tau_k})\boldsymbol{X}_i^*\right\|^{2+2\nu} < (pm)^{2+\nu}\Delta$, term (e8) is

$$\begin{aligned}
&\frac{1}{nm}\text{Vec}\Big(\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{i\tau_k})\boldsymbol{X}_i^*(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})(\widehat{\boldsymbol{\beta}}_{\tau_j}-\boldsymbol{\beta}_{\tau_j})^{\mathsf{T}}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{i\tau_j})\boldsymbol{X}_i^*\Big)\\
&=\frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_j}(\widehat{\boldsymbol{\varepsilon}}^*_{i\tau_j})\boldsymbol{X}_i^*\otimes\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\widehat{\boldsymbol{\varepsilon}}^*_{i\tau_k})\boldsymbol{X}_i^*\,\text{Vec}[(\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{\beta}_\tau)(\widehat{\boldsymbol{\beta}}_{\tau_k}-\boldsymbol{\beta}_{\tau_k})^{\mathsf{T}}]\\
&=\frac{1}{nm}O_p(1)O_p(1).
\end{aligned}$$

We have just shown that

$$\frac{1}{nm}\sum_{i=1}^{n}\widehat{\boldsymbol{X}}_i^{*\mathsf{T}}\widehat{\boldsymbol{\Sigma}}^*_{i\tau_k\tau_j}\widehat{\boldsymbol{X}}_i^* - \frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{i\tau_k})\boldsymbol{\varepsilon}^*_{i\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{i\tau_j}\boldsymbol{\Psi}_{\tau_j}(\boldsymbol{\varepsilon}^*_{i\tau_j})\boldsymbol{X}_i^* \xrightarrow{p} \mathbf{0}.$$

Application of the Markov LLN also yields

$$\frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}^*_{i\tau_k})\boldsymbol{\varepsilon}^*_{i\tau_k}\boldsymbol{\varepsilon}^{*\,\mathsf{T}}_{i\tau_j}\boldsymbol{\Psi}_{\tau_j}(\boldsymbol{\varepsilon}^*_{i\tau_j})\boldsymbol{X}_i^* - \frac{1}{nm}\sum_{i=1}^{n}\boldsymbol{X}_i^{*\mathsf{T}}\boldsymbol{\Sigma}^*_{i\tau_k\tau_j}\boldsymbol{X}_i^* \xrightarrow{p} \mathbf{0}.$$

Thus, application of the triangular inequality shows that $\widehat{\boldsymbol{D}}_0(\boldsymbol{\tau}) \xrightarrow{p} \boldsymbol{D}_0(\boldsymbol{\tau})$. □

Proof of Corollary 3.3.3.1.

Proof of **Corollary** 3.3.3.1 follows immediately from the proof of **Theorem** 3.3.3.

□

CHAPITRE IV

TROISIÈME ARTICLE : PENALIZED WEIGHTED ASYMMETRIC LEAST SQUARES REGRESSION FOR LONGITUDINAL DATA WITH FIXED-EFFECTS

Amadou Barry, Karim Oualkacha, Arthur Charpentier

Abstract. In this paper, we study the penalized weighted asymmetric least squares regression (expectile regression) for longitudinal data with fixed-effects. We use the l_1-penalty to shrink the fixed-effects incidental parameter and provide a sparse solution. The penalty shrinks the incidental parameter towards zero in order to provide accurate estimates of the covariates of interest. In addition, the l_1-penalty allows inference of time-invariant covariates. We apply a Bayesian information criterion to select the optimal shrinkage parameter. We propose a block-relaxation algorithm combined with a coordinate descent algorithm to compute efficiently the penalized estimator and the incidental parameter. The exhaustive simulations show that the proposed penalized estimator reduces bias and improves efficiency. The simulation results displayed its favorable qualities under various scenarios and its advantages in relation to existing methods. The usefulness of the penalized estimator is illustrated through analysis of an econometric panel dataset.

Keywords : Expectile regression, Quantile regression, Fixed-effects, Random-effects.

4.1 Introduction

The fixed-effects (FE) model is a commonly used model, particularly in econometrics, for the analysis of longitudinal data. The FE model in its specification takes into account the covariates that are not or cannot be measured and included in the model equation. In addition, its specification incorporates implicitly the possibility of a correlation between these omitted covariates and those present in the model. This latter characteristic makes the FE model an attractive model for longitudinal data analysis (Greene, 2012; Wooldridge, 2002; Schempf et Kaufman, 2012; Gardiner *et al.*, 2009). For example in the estimation of the effect of returns on schooling on the wage structure (response variable), the level of education (covariate) is correlated to the individual ability, which is often not included in the model. The same situation occurs in epidemiological and biological studies where the genetic (hidden relatedness or family structure) or environmental factors of the subjects is not measured when it is correlated with the covariate of interest of the model.

In the estimation process of the FE model parameters, we are faced with the presence of the infinite dimension of the individual fixed-effects, which is described as the incidental parameter problem (Lancaster, 2000; Neyman et Scott, 1948). This problem is solved for example, by application of the within-transformation strategy. The within-transformation strategy, which eliminates the individual-fixed-effects parameter by demeaning the different covariates and applying the ordinary least squares (OLS) to the transformed data. However, while eliminating the individual fixed-effects parameter, the within-transformation strategy also eliminates all time-invariant covariates present in the model. As a result, the coefficients of the time-invariant covariates are not estimated. This limitation can partly explain the small prevalence application of

the FE model in health science, where most often the treatment covariate is time-invariant (Dieleman et Templin, 2014).

Traditionally, the FE model estimates the effect of the covariates on the mean of the response variable. Recently, Barry *et al.* (2018b) applied the expectile regression (ER) in the FE framework to estimate the effects of the covariates beyond the conditional mean of the response variable. The ER with fixed-effects (ERFE) captures the influence of the covariates on the location, scale and shape of the distribution of the response variable. Thus, in addition to inheriting the attractive properties of the FE model, the ERFE model takes into account the heterogeneity of the effects of the covariates and unobserved heterogeneity. The authors presented an iterative within-transformation method to estimate the parameters of the ERFE model. They showed that this estimator is the generalization of the within-transformation estimator. As with the within-transformation method, the iterative within-transformation ERFE estimator uses a projection matrix to eliminate the incidental parameter from the model and apply the ER to the transformed data. The asymptotic properties of this new estimator and an estimator of its variance-covariance matrix are presented in (Barry *et al.*, 2018b). However, the iterative within-transformation method has the same drawback as the within-transformation strategy used to estimate the classical FE model. Namely, in the transformation process of the data, the time-invariant covariates are removed from the model and their coefficients are not estimated.

In this paper, we propose a new approach to estimate the parameters of both time-varying and time-invariant covariates of the ERFE model while mitigating the incidental parameter problem. This approach was first proposed in the quantile regression framework by Koenker (2004). Inspired by the characterization of the random effects

estimator as a penalized least squares estimator, Koenker applied the l_1-penalty in the quantile regression FE model to shrink the individual fixed-effects parameter toward zero. However, quantile regression combined with the l_1-penalty raise numerical problems since the corresponding objective function is not differentiable at zero. Because of that l_1-penalized quantile regression, unlike l_1-penalized expectile regression, is often computationally intensive, particularly in high dimensions (Mkhadri *et al.*, 2017).

Following Koenker (2004), we added a penalty to the ERFE model to constrain some entries of the incidental parameter vector toward zero. The penalty has the advantage of controlling the variability introduced by the incidental parameter and at the same time reducing the number of non-zero parameters to be estimated.

The application of a penalty shrinks the incidental parameter and prevents overfitting. There are two common penalties, l_1 and l_2, that are extensively used in the literature, (Friedman *et al.*, 2010). Both penalties are useful but their impact is quite different in practice. We use the penalty mainly to reduce the number of non-zero parameters of the nuisance parameter vector (incidental parameter) and for this purpose it is convenient to consider the l_1 penalty. The l_1 penalty shrinks the parameter toward zero and provides sparse solutions. The degree of shrinkage is controlled by the regularization parameter whose optimal value is hard to find. We rely on the Bayesian information criterion (BIC) in the selection of the best regularization parameter. In addition, we resort to special treatment (Mkhadri *et al.*, 2017) to compute the solution path for a grid of values of the regularization parameter to increase the speed of our algorithm.

What follows is structured this order : We present the model and estimation method in Section 4.2 and the asymptotic properties of the proposed estimator in Section 4.3. The simulation performance of the estimators is presented in Section 4.4 and the real data application of the method in Section 4.5. The conclusion is presented in Section 4.6. All proofs are available in the Appendix.

4.2 Models and Methods

4.2.1 Expectile and expectile regression

Expectile function is gradually becoming a broadly used statistic. The expectile of level $\tau \in [0,1]$ of a r.v Y is defined as

$$\mu_\tau(Y) = \underset{\theta \in \mathbb{R}}{\operatorname{argmin}}\, \mathbb{E}\{\rho_\tau(Y-\theta)\},$$

where $\rho_\tau(t) = |\tau - \mathbb{1}(t \leq 0)| \cdot t^2$ is the asymmetric square loss function that assigns weights τ and $1-\tau$ to positive and negative deviations respectively.

Given a sample, $\{(y_i)\}_{i=1}^n$, the corresponding empirical expectile of level τ can be defined as

$$\widehat{\mu}_\tau = \sum_{i=1}^{n} \frac{\psi_\tau(y_i - \widehat{\mu}_\tau)}{\sum_{l=1}^{n} \psi_\tau(y_l - \widehat{\mu}_\tau)} y_i, \tag{4.1}$$

where $\psi_\tau(t) = |\tau - \mathbb{1}(t \leq 0)|$ is the check function. Thus, expectile function is merely a weighted mean. The only subtlety is that the weights are not specified *a priory* but depend on the observed sample, (Kneib, 2013b).

Under the linear regression framework, Newey et Powell (1987) introduced the expectile regression (ER) model to study the effects of the covariates beyond the condi-

tional mean. The ER model offers a global view of the statistical landscape. Consider the linear regression model

$$y_i = \boldsymbol{x}_i{}^{\mathsf{T}}\boldsymbol{\beta} + \varepsilon_i \quad \text{with} \quad \mu_\tau(\varepsilon_i) = 0, \tag{4.2}$$

where $\boldsymbol{x}_i$ is a $p \times 1$ vector of covariates, y_i and ε_i are respectively the response variable and the random error with unspecified distribution function. $\boldsymbol{\beta} \in \mathbb{R}^p$ is the unknown parameters that need to be estimated. The assumption $\mu_\tau(\varepsilon_i) = 0$ ensures that the random error is centered on the τ-th expectile. In the linear regression settings, the ER model, for a fixed $\tau \in (0, 1)$, is given as

$$\mu_\tau(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i{}^{\mathsf{T}}\boldsymbol{\beta}_\tau. \tag{4.3}$$

The ER model estimates the effect of the covariates on the location, scale and shape of the distribution of the response variable. The ER estimator $\widehat{\boldsymbol{\beta}}_\tau$ can be derived by minimizing

$$\sum_{i=1}^{n} \rho_\tau\Big(y_i - \boldsymbol{x}_i{}^{\mathsf{T}}\boldsymbol{\beta}_\tau\Big)$$

over $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$ and as the asymmetric loss function $\rho_\tau(\cdot)$ is continuously differentiable, we have

$$\widehat{\boldsymbol{\beta}}_\tau = \left[\sum_{i=1}^{n} \psi_\tau(y_i - \boldsymbol{x}_i{}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_\tau)\boldsymbol{x}_i\boldsymbol{x}_i{}^{\mathsf{T}}\right]^{-1} \sum_{i=1}^{n} \psi_\tau(y_i - \boldsymbol{x}_i{}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_\tau)\boldsymbol{x}_i y_i. \tag{4.4}$$

The ER estimator is computed iteratively using, for example, the iterated re-weighted least squares algorithm.

Notice that the ER estimator $\widehat{\boldsymbol{\beta}}_{0.5}$ of level $\tau = 0.5$ corresponds to the OLS regression estimator. This makes the ER model a natural complement to the OLS regression.

4.2.2 Motivation of the penalty method

Before introducing our model, we briefly review an intriguing approach for the derivation of the random effect estimator, which motivates our proposed estimation method. Consider the standard linear model for longitudinal data

$$y_{ij} = \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta} + \alpha_i + \varepsilon_{ij}, \tag{4.5}$$

where y_{ij} is the response variable, $\boldsymbol{x}_{ij} = (x_{ij}^1, x_{ij}^2, \ldots, x_{ij}^p)^{\mathsf{T}} \in \mathbb{R}^p$ is the vector of covariates measured on individual i at time j, α_i is the individual parameter (which could be random or fixed), ε_{ij} is a random error and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the true parameter vector that need to be estimated. The matrix notation of model (4.5) is given as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{4.6}$$

where $\boldsymbol{y}$ is $nm \times 1$ vector, $\boldsymbol{X}$ is the $nm \times p$ matrix, $\boldsymbol{\alpha}$ and $\boldsymbol{\varepsilon}$ are respectively the $n \times 1$ and the $nm \times 1$ error vectors. The $nm \times n$ matrix $\boldsymbol{Z}$ is known as the incidence matrix and serves to identify the n distinct individuals of the sample.

Under the random-effects (RE) framework, $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$ being independent Gaussian vectors, Henderson (1950) proposed an estimator of the parameter of the model (4.6) using the joint maximum likelihood approach. The derived Gaussian random effects estimator which is known to be the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ is given as

$$\widehat{\boldsymbol{\beta}} = [\boldsymbol{X}^{\mathsf{T}}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} + \boldsymbol{Z}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\boldsymbol{Z}^{\mathsf{T}})^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}^{\mathsf{T}}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} + \boldsymbol{Z}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\boldsymbol{Z}^{\mathsf{T}})^{-1}\boldsymbol{y}. \tag{4.7}$$

The normal equations as well as all the calculation steps that give the expression (4.7) are presented in (Koenker, 2004; Robinson, 1991). The estimator can be considered as a Bayesian estimator or as a penalized least squares estimator (Ridge-type

estimator) (Robinson, 1991). This latter comparison is suggestive of a larger class of estimators that can be derived from the penalization of the individual parameter. Indeed, Koenker (2004) shows that the Gaussian random effects estimator minimizes the following objective function

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}\|^2_{\boldsymbol{\Sigma}_\varepsilon^{-1}} + \|\boldsymbol{\alpha}\|^2_{\boldsymbol{\Sigma}_\alpha^{-1}}.$$

We transpose this idea of penalization in the expectile regression with fixed-effects framework to mitigate the well-known incidental parameter problem. In the next section, we present the penalized expectile regression (PER) model for longitudinal data.

4.2.3 PER model for longitudinal data

Consider the following expectile regression model for longitudinal data

$$\mu_\tau(y_{ij}|\alpha_i, \boldsymbol{x}_{ij}) = \boldsymbol{x}_{ij}{}^\top \boldsymbol{\beta}_\tau + \alpha_i, \tag{4.8}$$

which is the analogous version to the linear model (4.5) with fixed-effects. We assume that $\mu_\tau(\varepsilon_{ij}) = 0$, to ensure that the random error is centered on the τ-th expectile. The parameter $\boldsymbol{\beta}_\tau$ shows the influence of the covariates on the location, scale and shape of the conditional distribution of the response variable. In this setting (4.8), only the parameter of interest is allowed to depend upon the parameter τ. We restrict the parameter α_i to be independent of τ because in most applications the number of observations per subject is relatively small compared to the number of individuals.

In addition to the estimation of the parameter of interest $\boldsymbol{\beta}_\tau$, the individual fixed-effects parameter α_i has to be estimated. But the infinite dimensional nature of the

parameter α_i brings some complications to the estimation process. In the classical fixed-effects model, these complications are bypassed by means of such strategies as the within-transformation strategy. The within-transformation strategy consists of subtracting from the equation (4.5) its time demeaned counterpart equation to eliminate the incidental parameter and adjust the ordinary least squares (OLS) method to the transformed data. Barry *et al.* (2018b) have applied this transformation strategy in the expectile regression framework with the fixed-effects model. They introduced the iterative within-transformation expectile regression with fixed-effects (ERFE) estimator. Details regarding the development of the ERFE model can be found here (Barry *et al.*, 2018b).

As with the FE model, the ERFE model elegantly solves the incidental parameter problem in the ER framework. But at the same time, the ERFE model has the same shortcoming, which is the non-identification of the effects of the time-invariant covariates.

In this paper, following the characterization of the RE estimator as a penalized estimator, we propose the penalized expectile regression with fixed-effects (PERFE) model.

For a sequence of expectiles $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)^\mathsf{T}$ and a vector of weight $\boldsymbol{v} = (v_1, \ldots, v_q)^\mathsf{T}$, the PERFE estimator is defined as the vector which minimizes the following objective function

$$R_{\lambda nm}(\boldsymbol{\beta_\tau}, \boldsymbol{\alpha}) = \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \rho_{\tau_k}\Big(y_{ij} - \boldsymbol{x_{ij}}^\mathsf{T}\boldsymbol{\beta}_{\tau_k} - \alpha_i\Big) + \lambda\sum_{i=1}^{n}|\alpha_i|. \qquad (4.9)$$

The vector of weights $\boldsymbol{v}$ is for controlling the relative influence of the q expectiles and $\lambda > 0$ is the tuning parameter. Notice that, when $\tau = 0.5$ the PERFE model corresponds to the Lasso regression introduced by Tibshirani (1996).

The PERFE model adds a penalty to the ERFE model. The penalty is applied to the individual fixed-effects parameter and is intended to solve the incidental parameter problem. In this way, the PERFE model estimates the parameter of interest of model (4.8) and at the same time mitigates the incidental parameter problem by providing a sparse solution. In addition, unlike the ERFE model, the PERFE model allows inference of time-invariant covariates.

There are two common penalties that are extensively used in the literature : the l_1-penalty and the l_2 penalty (Friedman *et al.*, 2010). We retain the l_1-penalty over the l_2 penalty for its statistical advantages related to our model (Tibshirani, 1996; Hastie *et al.*, 2001). The l_1-penalty has the ability to shrink the incidence parameter toward zero, provides sparse solutions for the estimation of $\boldsymbol{\alpha}$ while at the same time improving the performance of the estimate of the parameter of interest (Koenker, 2004; Tibshirani, 1996).

In the following, we present the algorithm generating the PERFE estimator. Since the objective function $R_{\lambda nm}$ is not differentiable at zero with respect to $\boldsymbol{\alpha}$, we cannot apply traditional algorithms such as the Gauss-Newton algorithm.

4.2.4 Algorithm and implementation

In this section, we use the block relaxation algorithm (de Leeuw, 1994) to efficiently solve the objective function of the PERFE model. The block relaxation divides the parameters into disjoint blocks and cycles through the blocks, updating only those parameters within the pertinent block at each stage of a cycle (Lange, 2013). In our case, we have two disjoint blocks : the block of the parameter of interest $\boldsymbol{\beta}_{\boldsymbol{\tau}}$ and the block of the nuisance parameter $\boldsymbol{\alpha}$.

Now, let $\widetilde{\boldsymbol{\alpha}}$ and $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{\tau}}$ be some current values of the parameters. We start the block relaxation algorithm with the block of the parameter of interest $\boldsymbol{\beta}_{\boldsymbol{\tau}}$. Define, for a fixed τ_k, the current residue $r_{ij\tau_k} = y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}}\widetilde{\boldsymbol{\beta}_{\tau_k}} - \widetilde{\alpha}_i$. Taking $\{\widetilde{\boldsymbol{\beta}}_{\boldsymbol{\tau}}, \widetilde{\boldsymbol{\alpha}}\}$ as initial values and assuming $\boldsymbol{\tau}$ and λ fixed, we update $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{\tau}}^{\text{new}}$ by solving :

$$\widetilde{\boldsymbol{\beta}}_{\boldsymbol{\tau}}^{\text{new}} = \underset{\boldsymbol{\beta}_{\boldsymbol{\tau}}}{\operatorname{argmin}} \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \psi_{\tau_k}(r_{ij\tau_k})\Big[r_{ij\tau_k} + \boldsymbol{x}_{ij}{}^{\mathsf{T}}(\widetilde{\boldsymbol{\beta}_{\tau_k}} - \boldsymbol{\beta}_{\tau_k})\Big]^2. \tag{4.10}$$

The implicit objective function in (4.10) is continuously differentiable with respect to $\boldsymbol{\beta}_{\boldsymbol{\tau}}$ and the new estimate can be computed iteratively using the iterated re-weighted least squares algorithm. Indeed, the new update of the PERFE estimator is the solution of

$$\widetilde{\boldsymbol{\beta}}_{\boldsymbol{\tau}}^{\text{new}} = \widetilde{\boldsymbol{\beta}}_{\boldsymbol{\tau}} + \Big[(\boldsymbol{V}\otimes\boldsymbol{X})^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{r}_{\boldsymbol{\tau}})(\mathbf{I}_q\otimes\boldsymbol{X})\Big]^{-1}(\boldsymbol{V}\otimes\boldsymbol{X})^{\mathsf{T}}\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{r}_{\boldsymbol{\tau}})\mathbb{1}_q\otimes\boldsymbol{r}_{\boldsymbol{\tau}}, \tag{4.11}$$

where $\boldsymbol{r}_{\boldsymbol{\tau}} = (\boldsymbol{r}_{\tau_1},\ldots,\boldsymbol{r}_{\tau_q})^{\mathsf{T}}$ and $\boldsymbol{r}_{\tau_k} = (r_{11\tau_k},\ldots,r_{nm\tau_k})^{\mathsf{T}}$. Notice that, the solution of each $\widetilde{\boldsymbol{\beta}}_{\tau_k}^{\text{new}}$, $k = 1,\ldots,q$, can be computed separately with

$$\widetilde{\boldsymbol{\beta}}_{\tau_k}^{\text{new}} = \widetilde{\boldsymbol{\beta}}_{\tau_k} + \Big[\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{r}_{\tau_k})\boldsymbol{X}\Big]^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{r}_{\tau_k})\boldsymbol{r}_{\tau_k}.$$

The second block to update is the nuisance parameter $\boldsymbol{\alpha}$ which is penalized by the non-differentiable Lasso penalty. In our algorithm, we implement the efficient coordinate descent algorithm to overcome the non-differentiability problem. The coordinate descent algorithm minimizes the objective function coordinate-wise by cycling over the parameter space sequentially instead of determining the minimum simultaneously. The coordinate descent algorithm is an algorithm that has proven itself in the estimation of parameters in high dimension (Friedman *et al.*, 2010). The algorithm is simple, efficient, fast and stable, (Wu et Lange, 2008). The coordinate descent algorithm is convergent for an objective function like ours : objective function composed of a convex, differentiable function and a separable function, (Tseng, 2001).

The coordinate descent algorithm updates each component α_i by solving :

$$\underset{\alpha_i}{\operatorname{argmin}} \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \psi_{\tau_k}(r_{ij\tau_k})[r_{ij\tau_k} + (\widetilde{\alpha}_i - \alpha_i)]^2 + \lambda \sum_{i=1}^{n} |\alpha_i|, \; i = 1, \ldots, n. \quad (4.12)$$

We compute the gradient of the objective function (4.12) with respect to α_i, which only exists if $\alpha_i \neq 0$. When $\alpha_i > 0$, the gradient is

$$-2\sum_{k=1}^{q}\sum_{j=1}^{m} v_k \psi_{\tau_k}(r_{ij\tau_k})[r_{ij\tau_k} + (\widetilde{\alpha}_i - \alpha_i)] + \lambda.$$

A similar expression exists when $\alpha_i < 0$, and $\alpha_i = 0$ is treated separately. After some calculations we show that the coordinate-wise update is a function of the soft-thresholding operator

$$\widetilde{\alpha}_i^{\text{new}} = \frac{S\Big\{\widetilde{\alpha}_i m^{-1} \sum_{k=1}^{q}\sum_{j=1}^{m} v_k \psi_{\tau_k}(r_{ij\tau_k}) + m^{-1} \sum_{k=1}^{q}\sum_{j=1}^{m} v_k \psi_{\tau_k}(r_{ij\tau_k})(r_{ij\tau_k}), \lambda\Big\}}{m^{-1} \sum_{k=1}^{q}\sum_{j=1}^{m} v_k \psi_{\tau_k}(r_{ij\tau_k})}$$

where $S(z, u) = (|z| - u)_+ \operatorname{sign}(z)$ is the soft-thresholding operator.

Finally, the block relaxation algorithm for the PERFE model with the Lasso penalty is detailed in **Algorithm 1**.

Algorithm 1 : The PERFE algorithm with the Lasso penalty

- Initialize $(\widetilde{\boldsymbol{\beta}}_{\boldsymbol{\tau}}, \widetilde{\boldsymbol{\alpha}})$.
- Iterate (1) and (2) until convergence :
  1. Update $\widetilde{\boldsymbol{\beta}}_{\boldsymbol{\tau}}$ with an iterated re-weighted least squares algorithm. For $k = 1, \ldots, q$,
     (a) Compute $\boldsymbol{r}_{\tau_k} = \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{Z}\widetilde{\boldsymbol{\alpha}}$.
     (b) Compute $\widetilde{\boldsymbol{\beta}}_{\tau_k}^{\text{new}} = \widetilde{\boldsymbol{\beta}}_{\tau_k} + \left[\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{r}_{\tau_k})\boldsymbol{X}\right]^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{r}_{\tau_k})\boldsymbol{r}_{\tau_k}$.
     (c) Set $\widetilde{\boldsymbol{\beta}}_{\tau_k} = \widetilde{\boldsymbol{\beta}}_{\tau_k}^{\text{new}}$.
  2. Update $\widetilde{\boldsymbol{\alpha}}$ with a cyclic coordinate descent algorithm. For $i = 1, \ldots, n$ and $k = 1, \ldots, q$,
     (a) Re-compute $r_{ij\tau_k} = y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}}\widetilde{\boldsymbol{\beta}}_{\tau_k} - \widetilde{\alpha}_i$.
     (b) Compute

$$\widetilde{\alpha}_i^{\text{new}} = \frac{S\left\{\widetilde{\alpha}_i m^{-1}\sum_{k=1}^{q}\sum_{j=1}^{m} v_k\psi_{\tau_k}(r_{ij\tau_k}) + m^{-1}\sum_{k=1}^{q}\sum_{j=1}^{m} v_k\psi_{\tau_k}(r_{ij\tau_k})(r_{ij\tau_k}), \lambda\right\}}{m^{-1}\sum_{k=1}^{q}\sum_{j=1}^{m} v_k\psi_{\tau_k}(r_{ij\tau_k})}.$$

     (c) Set $\widetilde{\alpha}_i = \widetilde{\alpha}_i^{\text{new}}$.

Usually, in the penalty framework, we standardize the covariates before applying the algorithm. This ensures all covariates are treated equally in the regularization process (Hastie *et al.*, 2001). In our case, the covariates are not penalized, the penalty is applied to the incidental parameter and its columns are already in the same scale

and have same units. So it is not necessary to standardize.

Finding the effective value of the regularization parameter, among all its possible values, can be tedious and time-consuming. In order to limit the search time, we resort to special treatment, similar to those used by Mkhadri *et al.* (2017), to compute the solutions for a fine grid of λ. The fine grid is obtain by calculating the $\lambda_{\max}$ which is the smallest λ to set all $\alpha_i,\ 1 \leq i \leq n$, to zero. We start the process by deriving the estimator $\widehat{\boldsymbol{\beta}}_{\tau_k}(\boldsymbol{\alpha} = \mathbf{0}) = \widehat{\boldsymbol{\beta}}_{\tau_k}^{\boldsymbol{\alpha_0}},\ k = 1, \ldots, q$, without any fixed-effect, $\boldsymbol{\alpha} = \mathbf{0}$:

$$\widehat{\boldsymbol{\beta}}_{\tau_k}^{\boldsymbol{\alpha_0}} = \underset{\boldsymbol{\beta}_{\tau_k}}{\operatorname{argmin}} \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho_{\tau_k}(y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}} \boldsymbol{\beta}_{\tau_k}).$$

Then using the Karush-Khun-Tucker (KKT) conditions

$$\frac{1}{m} \left| \sum_{k=1}^{q} \sum_{j=1}^{m} v_k \psi_{\tau_k}(y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}} \widehat{\boldsymbol{\beta}}_{\tau_k}^{\boldsymbol{\alpha_0}})(y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}} \widehat{\boldsymbol{\beta}}_{\tau_k}^{\boldsymbol{\alpha_0}}) \right| \leq \lambda, \text{ for } i \in \{1, \ldots, n\}, \quad (4.13)$$

we have

$$\lambda_{\max} = \frac{1}{m} \max_{i \leq n} \left| \sum_{k=1}^{q} \sum_{j=1}^{m} v_k \psi_{\tau_k}(y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}} \widehat{\boldsymbol{\beta}}_{\tau_k}^{\boldsymbol{\alpha_0}})(y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}} \widehat{\boldsymbol{\beta}}_{\tau_k}^{\boldsymbol{\alpha_0}}) \right|, \quad (4.14)$$

where $\psi_{\tau_k}(t) = |\tau_k - \mathbb{1}(t < 0)|$ is the check function. We put $\lambda_{\min} = 10^{-4} \lambda_{\max}$ and we place K points uniformly in the log-scale between $\lambda_{\min}$ and $\lambda_{\max}$. In our simulation study we set $K = 10$.

We illustrate the special treatment implemented to reduce the time-consumption in the search for the best regularization parameter. Consider the sample generated according to the following simulation model :

$$y_{ij} = x_{ij} + \alpha_i + \varepsilon_{ij},\ i = 1, \ldots, 50;\ j = 1, \ldots, 10$$

where $x_{ij} = \alpha_i + u_{ij}$. The variables α_i, u_{ij}, ε_{ij} are generated from a standard normal distribution $\mathcal{N}(0,1)$ and $\boldsymbol{\alpha} = (\underbrace{0.5,\ldots,0.5}_{5 \text{ times}},\underbrace{0,\ldots,0}_{45 \text{ times}})$ is a sparse vector with 5 non-zero values. **Figure 4.1** presents the estimated fixed-effects, $\widehat{\boldsymbol{\alpha}}$, as a function of the index of the shrinkage parameter, λ. The non-zero fixed-effects values are in red and the blue lines identify the best lambda according to $\text{BIC}_{1\lambda}$ and $\text{BIC}_{2\lambda}$ criteria, defined below. At the beginning, left side of λ-axis, there is no shrinkage at all. All fixed-effects are different from zero. Then, as the amount of shrinkage gradually increases, from left to right of λ-axis, we see all the fixed-effects converging toward zero. $\text{BIC}_{2\lambda}$ criterion is more likely to choose a sparse model with active fixed-effects parameters.

Arbitrary choice of the regularization parameter affects the inference of the parameter of interest $\boldsymbol{\beta}_\tau$. Large value of $\lambda \to \infty$ shrinks all the individual fixed-effects parameters to zero and corresponds to a particular case of the model proposed by Barry *et al.* (2018a). While a small value of $\lambda \cong 0$ does not shrink them at all and is equivalent to the ERFE model of Barry *et al.* (2018b). Beyond these particular values, the selection of the best tuning parameter is an open problem. However, there are strategies developed to circumvent the problem. This is discussed next.

4.2.5 Choice of the regularization parameter

In high dimension the optimal value of the regularization parameter can be determined by a cross-validation method. The cross-validation method consists of partitioning the dataset into two classes : the training dataset with which the model is adjusted and the test dataset with which the predictive power of the model is evaluated. The details of this approach can be found in the excellent book of Has-

tie *et al.* (2001). However, the efficiency of the cross-validation method requires the availability of large amounts of data. For this reason, other criteria such as generalized cross-validation (GCV), Akaike information criterion (AIC) or Bayes information criterion (BIC) are often recommended. The BIC is particularly appreciated, among other things, for its consistency (Galvao et Montes-Rojas, 2010).

In this section, we present two Bayesian criteria to guide us in the selection of the best regularized parameter. We start with the selection criterion, known as the least squares approximation (LSA), defined as

$$\mathrm{BIC}_{1\lambda} = (\widehat{\boldsymbol{\beta}}_\lambda - \widehat{\boldsymbol{\beta}}_0)^\intercal \widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\beta}}_0)(\widehat{\boldsymbol{\beta}}_\lambda - \widehat{\boldsymbol{\beta}}_0) + (nm)^{-1}\mathrm{df}_\lambda \log(nm) \tag{4.15}$$

and proposed by Wang et Leng (2007). The estimators $\widehat{\boldsymbol{\beta}}_\lambda$ and $\widehat{\boldsymbol{\beta}}_0$ are respectively the penalized and the un-penalized estimator and $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}_0) = \widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}_0)$. In our case we choose $\widehat{\boldsymbol{\beta}}_0$ and $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}_0)$ to be the ERFE estimator and its variance-covariance matrix proposed by Barry *et al.* (2018b). The term df_λ is a measure of the effective dimensionality of the fitted model.

Our second choice is the selection criterion proposed by Wang *et al.* (2007a,b) in the quantile regression framework. This criterion is defined as

$$\mathrm{BIC}_{2\lambda} = (nm)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho_\tau(y_{ij} - \boldsymbol{x}_{ij}{}^\intercal \widehat{\boldsymbol{\beta}}_{\tau\lambda} - \alpha_i) + \frac{1}{2}(nm)^{-1}\mathrm{df}_\lambda \log(nm) \tag{4.16}$$

where df_λ is a measure of the effective dimensionality of the fitted model. Following Galvao et Montes-Rojas (2010), we determine the generalized degrees of freedom df_λ for both criteria as the dimension of the set $\{\#\boldsymbol{\beta} \cup \{i|\ |\alpha_i| > \kappa\}\}$, where κ is a sufficiently small chosen number. We choose three different values of κ in our simulation study to evaluate the relative sensitivity of the criterion.

Figure 4.2 shows the values of the criteria, $BIC_{1\lambda}$ and $BIC_{2\lambda}$, according to the logarithm of the regularization parameter defined in the previous model. On the left panel **Figure 4.2**, we have the results of $BIC_{1\lambda}$ and on the right panel the results of $BIC_{2\lambda}$ and on the x-axis the values of $\log(\lambda)$. As mentioned above, both BICs suggested a sparse model. But, $BIC_{2\lambda}$ criterion suggest a sparse model with active fixed-effects parameters and is more accurate than $BIC_{1\lambda}$ criterion.

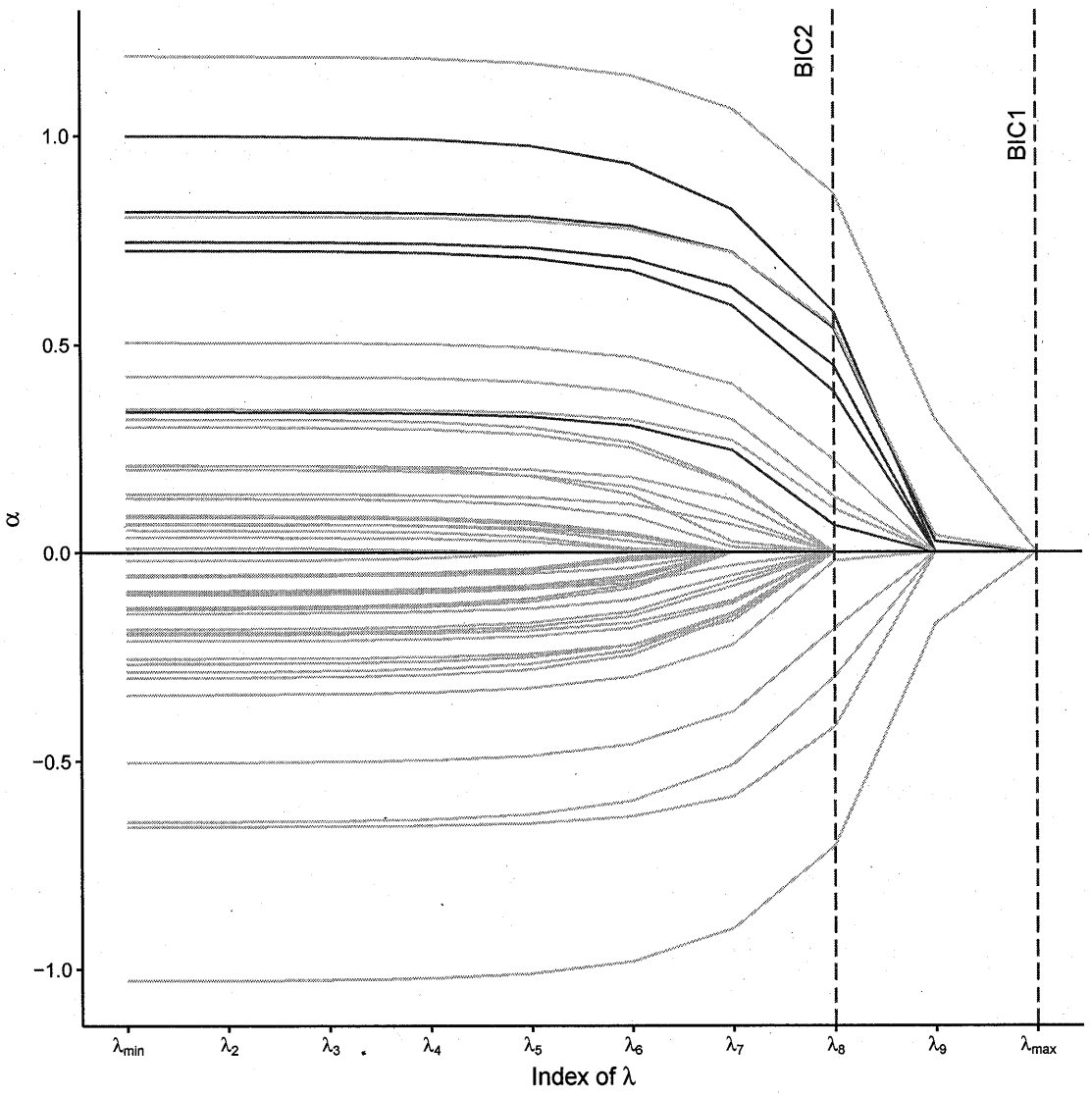

Figure 4.1: Path solutions of the fixed-effects parameter estimates, using l_1-Lasso penalty

4.3 Asymptotics

We derive the asymptotic properties of the PERFE estimator for $n, m \to \infty$. Consider the following risk function

$$R_{nmq}(\boldsymbol{\delta}, \lambda_m) = \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{j=1}^{m} v_k \Big[\rho_{\tau_k} \Big\{ y_{ij} - \mu_{ij\tau_k} - \boldsymbol{x}_{ij}{}^{\mathsf{T}} \boldsymbol{\delta}_{1\tau_k} / \sqrt{N} - \delta_{0i} / \sqrt{m} \Big\} - \rho_{\tau_k} \{ y_{ij} - \mu_{ij\tau_k} \} \Big] + \lambda_m \sum_{i=1}^{n} |\alpha_i - \delta_{0i} / \sqrt{m}| - |\alpha_i|, \tag{4.17}$$

where $\mu_{ij\tau_k} = \boldsymbol{x}_{ij}{}^{\mathsf{T}} \boldsymbol{\beta}_{\tau_k} + \alpha_i$ and $N = nm$. The tuning parameter $\lambda_m > 0$ shrinks the individual specific effect toward zero. The risk function R_{nmq} is convex and is minimized by $\widehat{\boldsymbol{\delta}} = (\widehat{\boldsymbol{\delta}}_0{}^{\mathsf{T}}, \widehat{\boldsymbol{\delta}}_{1\boldsymbol{\tau}}{}^{\mathsf{T}})^{\mathsf{T}}$, where

$$\widehat{\boldsymbol{\delta}}_0 = \begin{pmatrix} \widehat{\delta}_{01} \\ \vdots \\ \widehat{\delta}_{0n} \end{pmatrix} = \begin{pmatrix} \sqrt{m}(\widehat{\alpha}_1 - \alpha_1) \\ \vdots \\ \sqrt{m}(\widehat{\alpha}_n - \alpha_n) \end{pmatrix} \tag{4.18}$$

and

$$\widehat{\boldsymbol{\delta}}_{1\boldsymbol{\tau}} = \begin{pmatrix} \widehat{\boldsymbol{\delta}}_{1\tau_1} \\ \vdots \\ \widehat{\boldsymbol{\delta}}_{1\tau_q} \end{pmatrix} = \begin{pmatrix} \sqrt{N}(\widehat{\boldsymbol{\beta}}_{\tau_1} - \boldsymbol{\beta}_{\tau_1}) \\ \vdots \\ \sqrt{N}(\widehat{\boldsymbol{\beta}}_{\tau_q} - \boldsymbol{\beta}_{\tau_q}) \end{pmatrix}. \tag{4.19}$$

Assume the following regularity condition

A1. The data $\{(\boldsymbol{y}_i, \boldsymbol{X}_i)\}_{i=1}^n$ are independent across i, and

$$\text{Var}\left[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\right] = \mathbb{E}\left[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\boldsymbol{\varepsilon}_{i\tau}\boldsymbol{\varepsilon}_{i\tau}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau})\right] = \boldsymbol{\Sigma}_{i\tau}, \text{ where}$$

$$\boldsymbol{\varepsilon}_{i\tau} = (\varepsilon_{i1\tau}, \ldots, \varepsilon_{im\tau})^{\mathsf{T}}, \ \varepsilon_{ij\tau} = y_{ij} - \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta}_\tau \text{ and } \boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_{i\tau}) = \left[\text{diag}(\psi_\tau(\varepsilon_{ij\tau}))\right]_{j=1}^m.$$

A2. Assume the limiting forms of the following matrices are positive definite

$$\boldsymbol{D}_0(\boldsymbol{\tau}) = \lim_{\substack{m\to\infty \\ n\to\infty}} m^{-1} \begin{pmatrix} \boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{\tau}}\boldsymbol{v}\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{Z} & \boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{\tau}}\boldsymbol{V} \otimes \boldsymbol{Z}^{\mathsf{T}}\boldsymbol{X}/\sqrt{n} \\ \boldsymbol{V}\boldsymbol{\Sigma}_{\boldsymbol{\tau}}\boldsymbol{v} \otimes \boldsymbol{X}^{\mathsf{T}}\boldsymbol{Z}/\sqrt{n} & \boldsymbol{V}\boldsymbol{\Sigma}_{\boldsymbol{\tau}}\boldsymbol{V} \otimes \boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}/n \end{pmatrix}$$

$$\boldsymbol{D}_1(\boldsymbol{\tau}) = \lim_{\substack{m\to\infty \\ n\to\infty}} m^{-1} \begin{pmatrix} \sum_{k=1}^q v_k \boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_k}]\boldsymbol{Z} & v_1\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_1}]\boldsymbol{X}/\sqrt{n} & \cdots & v_q\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_q}]\boldsymbol{X}/\sqrt{n} \\ v_1\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_1}]\boldsymbol{Z}/\sqrt{n} & v_1\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_1}]\boldsymbol{X}/n & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ v_q\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_q}]\boldsymbol{Z}/\sqrt{n} & \boldsymbol{0} & \cdots & v_q\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_q}]\boldsymbol{X}/n \end{pmatrix}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\tau}} = \text{Var}[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{\boldsymbol{\tau}})\boldsymbol{\varepsilon}_{\boldsymbol{\tau}}]$ and $\boldsymbol{\Psi}_{\tau_k} = \boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{\tau_k})$.

A3. $\max_{i,j} \|\boldsymbol{x}_{ij}\| < M$.

A4. $\lambda_m/\sqrt{m} \to \lambda_0$, where λ_0 is some constant.

Condition **A1** ensures independence across individuals, but permits a within-dependency between observations of the same individual and allows heterogeneity across individuals. Condition **A2** is a full rank condition and is used to invoke the Lindeberg-Feller Central Limit Theorem. Condition **A3** is useful both for the application of the Lindeberg-Feller Central Limit Theorem and for ensuring the finite dimensional convergence of the objective function. Condition **A4** is from Knight et Fu (2000) who derived the limiting distribution of the class of Bridge estimators. We use the

same approach to derive the asymptotic distribution of the PERFE estimator. We start with a result on the limiting function of the R_{nmq} function.

Theorem 4.3.1. *Under conditions **A1-A4**, the first component $\widehat{\boldsymbol{\delta}}_1$ minimizing the objective function R_{nmq} has the same limiting distribution as the first component of the minimizer of*

$$R_{0q}(\boldsymbol{\delta}) = -2\boldsymbol{\delta}^T\boldsymbol{B} + \boldsymbol{\delta}^T\boldsymbol{D}_1(\boldsymbol{\tau})\boldsymbol{\delta} + \lambda_0\boldsymbol{\delta}^T\boldsymbol{s},$$

where $\boldsymbol{B}$ is a zero mean Gaussian vector with covariance matrix $\boldsymbol{D}_0(\boldsymbol{\tau})$, and $\boldsymbol{s} = (\boldsymbol{s}_0{}^T, \mathbf{0}_{pq}{}^T)^T$ and $\boldsymbol{s}_0 = (\text{sign}(\alpha_1), \ldots, \text{sign}(\alpha_n))^T$

The results of **Theorem 4.3.1** will lead to the derivation of the asymptotic property of the PERFE estimator. Assume that

A2'. The limiting forms of the following matrices are positive definite

$$\widetilde{\boldsymbol{D}}_{1q}(\boldsymbol{\tau}) = \lim_{\substack{m\to\infty \\ n\to\infty}} (\mathbf{I}_q \otimes \boldsymbol{X})^\mathsf{T}\mathbb{E}[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{\boldsymbol{\tau}})]\boldsymbol{M}_{q\boldsymbol{Z}}(\boldsymbol{\tau})(\boldsymbol{V}\otimes\boldsymbol{X})/nm$$

$$\widetilde{\boldsymbol{D}}_{0q}(\boldsymbol{\tau}) = \lim_{\substack{m\to\infty \\ n\to\infty}} (\boldsymbol{V} \otimes \boldsymbol{X})^\mathsf{T}\boldsymbol{M}_{q\boldsymbol{Z}}{}^\mathsf{T}(\boldsymbol{\tau})\boldsymbol{\Sigma}_{\boldsymbol{\tau}}\boldsymbol{M}_{q\boldsymbol{Z}}(\boldsymbol{\tau})(\boldsymbol{V}\otimes\boldsymbol{X})/nm.$$

Matrix $\boldsymbol{M}_{q\boldsymbol{Z}}(\boldsymbol{\tau}) = \mathbb{I}_{qnm} - \boldsymbol{P}_{q\boldsymbol{Z}}(\boldsymbol{\tau})$ is an idempotent matrix. Its complement $\boldsymbol{P}_{q\boldsymbol{Z}}(\boldsymbol{\tau})$ which is also idempotent is defined as

$$\boldsymbol{P}_{q\boldsymbol{Z}}(\boldsymbol{\tau}) = (\mathbb{1}_q \otimes \boldsymbol{Z})\Big[(\boldsymbol{v}\otimes\boldsymbol{Z})^\mathsf{T}\mathbb{E}[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{\boldsymbol{\tau}})](\mathbb{1}_q\otimes\boldsymbol{Z})\Big]^{-1}(\boldsymbol{v}\otimes\boldsymbol{Z})^\mathsf{T}\mathbb{E}[\boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{\boldsymbol{\tau}})].$$

These matrices are identical to those defined in the development of the ERFE model by Barry *et al.* (2018b). They are orthogonal to the time-invariant covariates and their presence in the expression of the variance covariance matrix does not allow the derivation of the variance of the latter. Thus, the PERFE model allows estimation of

the time-invariant covariates, but does not compute their variance. Other methods, like bootstrap, will be used to compute the variance of the time-invariant covariates.

Corollary 4.3.1.1. *Assume conditions* ***A1, A2', A3-A4*** *and* $\mathbb{E}|\psi_\tau(\varepsilon_{ij\tau})|^{4+\nu} < \Delta < \infty$ *and* $\mathbb{E}|\varepsilon_{ij\tau}|^{4+\nu} < \Delta < \infty$ *for some* $\nu > 0$. *Then*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} - \boldsymbol{\beta}_{\boldsymbol{\tau}}) \xrightarrow{d} \mathcal{N}\Big(Bias_{\boldsymbol{\tau}}(\lambda), \widetilde{\boldsymbol{D}}_{1q}^{-1}(\boldsymbol{\tau})\widetilde{\boldsymbol{D}}_{0q}(\boldsymbol{\tau})\widetilde{\boldsymbol{D}}_{1q}^{-1}(\boldsymbol{\tau})\Big),$$

where

$$Bias_{\boldsymbol{\tau}}(\lambda) = \frac{\lambda_0}{2m}\Big[(\mathbf{I}_q \otimes \boldsymbol{X})^T \mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{M}_{q\boldsymbol{Z}}(\boldsymbol{\tau})(\boldsymbol{V}\otimes\boldsymbol{X})\Big]^{-1}(\boldsymbol{V}\otimes\boldsymbol{X})^T\boldsymbol{P}_{q\boldsymbol{Z}}{}^T(\boldsymbol{\tau})(\boldsymbol{v}\otimes\boldsymbol{Z})\boldsymbol{s}_0.$$

A special case of **Corollary 4.3.1.1** is the result for one expectile which corresponds to setting the total number of expectiles q and weight $\boldsymbol{v}$ to 1. The result for one expectile is presented in **Corollary 4.3.1.2**. Let

$$\widetilde{\boldsymbol{D}}_1(\tau) = \lim_{\substack{m\to\infty \\ n\to\infty}} \boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{M}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}/nm,$$

$$\widetilde{\boldsymbol{D}}_0(\tau) = \lim_{\substack{m\to\infty \\ n\to\infty}} \boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}_{\boldsymbol{Z}}{}^{\mathsf{T}}(\tau)\boldsymbol{\Sigma}_\tau\boldsymbol{M}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}/nm,$$

where $\boldsymbol{M}_{\boldsymbol{Z}}(\tau) = \mathbb{I}_{nm} - \boldsymbol{P}_{\boldsymbol{Z}}(\tau)$ and

$$\boldsymbol{P}_{\boldsymbol{Z}}(\tau) = \boldsymbol{Z}\Big[\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{Z}\Big]^{-1}\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)].$$

Corollary 4.3.1.2. *Assume conditions* ***A1, A3-A4*** *and* $\mathbb{E}|\psi_\tau(\varepsilon_{ij\tau})|^{4+\nu} < \Delta < \infty$ *and* $\mathbb{E}|\varepsilon_{ij\tau}|^{4+\nu} < \Delta < \infty$ *for some* $\nu > 0$. *If the matrices* $\widetilde{\boldsymbol{D}}_1(\tau)$ *and* $\widetilde{\boldsymbol{D}}_0(\tau)$ *are positive definite then*

$$\sqrt{nm}(\widehat{\boldsymbol{\beta}}_{\tau\lambda} - \boldsymbol{\beta}_{\tau\lambda}) \xrightarrow{d} \mathcal{N}\Big(Bias_{\boldsymbol{\tau}}(\lambda), \widetilde{\boldsymbol{D}}_1^{-1}(\tau)\widetilde{\boldsymbol{D}}_0(\tau)\widetilde{\boldsymbol{D}}_1^{-1}(\tau)\Big)$$

where

$$Bias_{\boldsymbol{\tau}}(\lambda) = \frac{\lambda_0}{2m}\Big[\boldsymbol{X}^T\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{M}_{\boldsymbol{Z}}(\tau)\boldsymbol{X}\Big]^{-1}\boldsymbol{X}^T\boldsymbol{P}_{\boldsymbol{Z}}{}^T(\tau)\boldsymbol{Z}\boldsymbol{s}_0.$$

4.4 Simulations

In this section, the small sample performance of the PERFE estimators is evaluated through extensive simulation studies. The simulation design proposed here is replicated from Barry *et al.* (2018b). The random samples are generated from the following linear model

$$y_{ij} = \beta_0 + x_{ij}\beta_1 + \alpha_i + (1 + \gamma x_{ij})\varepsilon_{ij}, \tag{4.20}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Two versions of model (4.20) are considered with respect to the parameter $\gamma \in \{0, 1/10\}$. A location shift model (M_0) corresponding to $\gamma = 0$, which helps assess the performance of the estimators for a homoscedastic scenario and a location-scale shift model ($M_{1/10}$) corresponding to $\gamma = 1/10$ to assess the performance of the estimators in the presence of heteroscedasticity.

The individual-specific effect α_i and the disturbance ε_{ij} are generated by the same distribution in three different models : normal distribution $\mathcal{N}(0, 1)$, Student distribution t_3 with 3 degrees of freedom, and chi-squared distribution χ^2_3 with 3 degrees of freedom. In addition, for each distribution the random error is centred on its τ-th expectile, $\mu_\tau(\varepsilon_{ij})$. The continuous explicative covariate x is generated by a normal distribution in the location shift model and by a chi-squared distribution with 3 degrees of freedom in the location-scale shift model. The parameters $\beta_0 = \beta_1 = 0$ are set to zero. The number of subjects and the number of within-subject observations are respectively $n \in (50, 100, 250)$ and $m \in (5, 15, 25)$. The extensive simulation is carried out with 400 replications. In each case the focus is on the effect of the covariate (x_{ij}) at the expectiles $\tau \in \{0.25, 0.5, 0.75\}$.

We compute the solutions path for a grid of λ by placing $K = 10$ points uniformly in the log-scale between $\lambda_{\min} = 10^{-4}\lambda_{\max}$ and $\lambda_{\max}$. We estimated both criterion $\text{BIC}_{1\lambda}$ and $\text{BIC}_{2\lambda}$ presented in Section 4.2 and we rely on it to select the optimal value of the regularization parameter. We showed their performance for a single replication. We also estimated three different values of $\kappa \in \{0.001, 0.01, 0.1\}$ to assess the relative sensitivity of the criterion.

We also estimated the ERFE model of Barry *et al.* (2018b) and the penalized quantile regression with fixed-effects (PQRFE) model of Koenker (2004). The regularization parameter $\lambda = \sigma_\varepsilon/\sigma_\alpha$ is chosen for the PQRFE model, as suggested by Koenker (2004). Note that, this value is not necessarily optimal, especially in a context where the distribution is not Gaussian. The simulation of the PQRFE model was carried out by choosing, for each distribution, the asymmetric points for which the quantiles are equal to the expectiles. For example, the Gaussian quantiles of $\tau = (0.33, 0.5, 0.67)$ correspond to the Gaussian expectiles of $\tau = (0.25, 0.5, 0.75)$. The quality of the estimators is evaluated by the distribution of their bias represented as box-plots. All computations are performed with the R v3.4.0 statistical programming language (R Core Team, 2018). The PQRFE model is estimated with the rqpd v0.6 package (Koenker et Bache, 2014) for R.

Here, we show the results of the estimators adjusted to the data generated by the normal distribution. The results of the estimators for the other distributions, Student and Chi-square, are set out in the **Supplementary material I**.

Figure 4.3 and **Figure 4.4** present the results of the distribution of the bias for the location-shift (M_0) and the location-scale-shift ($M_{1/10}$) scenario. We have, from

left to the right of the x-axes, the ERFE, PER1, PER2 and PQRFE estimators. The estimators PER1 and PER2 are respectively the PERFE estimator corresponding to the optimal value of λ according to the criteria $\text{BIC}_{1\lambda}$ and $\text{BIC}_{2\lambda}$. In columns, we have the results according to the values of $\tau \in (\tau_1, \tau_2, \tau_3)$, and in rows, the results according to the values of the number of subjects n and the number of within-subject observation m : 50-5, 50-15, 250-5, 250-15.

Overall, we observe a relatively small bias for both estimators, ERFE, PER1, PER2 and PQRFE. The first thing of immediate interest is the decreasing of the range and the interquartile of the different box-plots as well as the number of extreme values as the size of the sample increases. This is in line with the asymptotic properties of the estimators and the fact that the fixed-effects is consistent when the number of within-observation goes to infinity (Hansen, 2007).

In both scenarios, (M_0) and $(M_{1/10})$, the optimal PERFE estimators (PER1, PER2) outperform the ERFE estimator. This result is consistent with the fact that the PER1 and PER2 estimators are the optimal estimators according to both criteria and that the ERFE estimator corresponds to λ_{min} which does not at all shrink the fixed-effects. In the location-shift scenario, (M_0), the expectile family estimators perform better than the PQRFE estimator. It is observed that the distribution of bias of the PQRFE estimator is less centered at zero compared to the other estimators. In the location-scale-shift scenario, $(M_{1/10})$, the PQRFE estimator performs slightly better than the PERFE estimators when the sample size is small. However, they have comparable performance as the sample size increases.

Similar conclusions are found with other distributions (Student and Chi-square) and

for both scenarios, (M_0) and $(M_{1/10})$: small bias for all estimators, decrease of the range and the interquartile as well as the number of extreme values as the size of the sample increases. That said, the PERFE estimators (PER1 and PER2) outperform the PQRFE estimator in both distributions (Student and Chi-square) and scenarios (M_0) and $(M_{1/10})$, see the results in the Appendix (**Supplementary material I**).

Figure 4.5 and **Figure 4.6** show the results of both criteria $\text{BIC}_{1\lambda}$ and $\text{BIC}_{2\lambda}$ for a single replication. The results of the three values of $\kappa = (0.001, 0.01, 0.1)$ are identical for both criteria, meaning that the two criteria do not depend on the variability of the values of the parameter κ. In the location-shift scenario (M_0), the results show that both criteria select a dense model, which correspond to a λ that does not shrink at all the fixed-effects, **Figure 4.5**. In the location-scale-shift scenario $(M_{1/10})$, the results show that both criteria choose a sparse model, **Figure 4.6**. The results of the other distributions (Student and Chi-square) are in the (**Supplementary material I**).

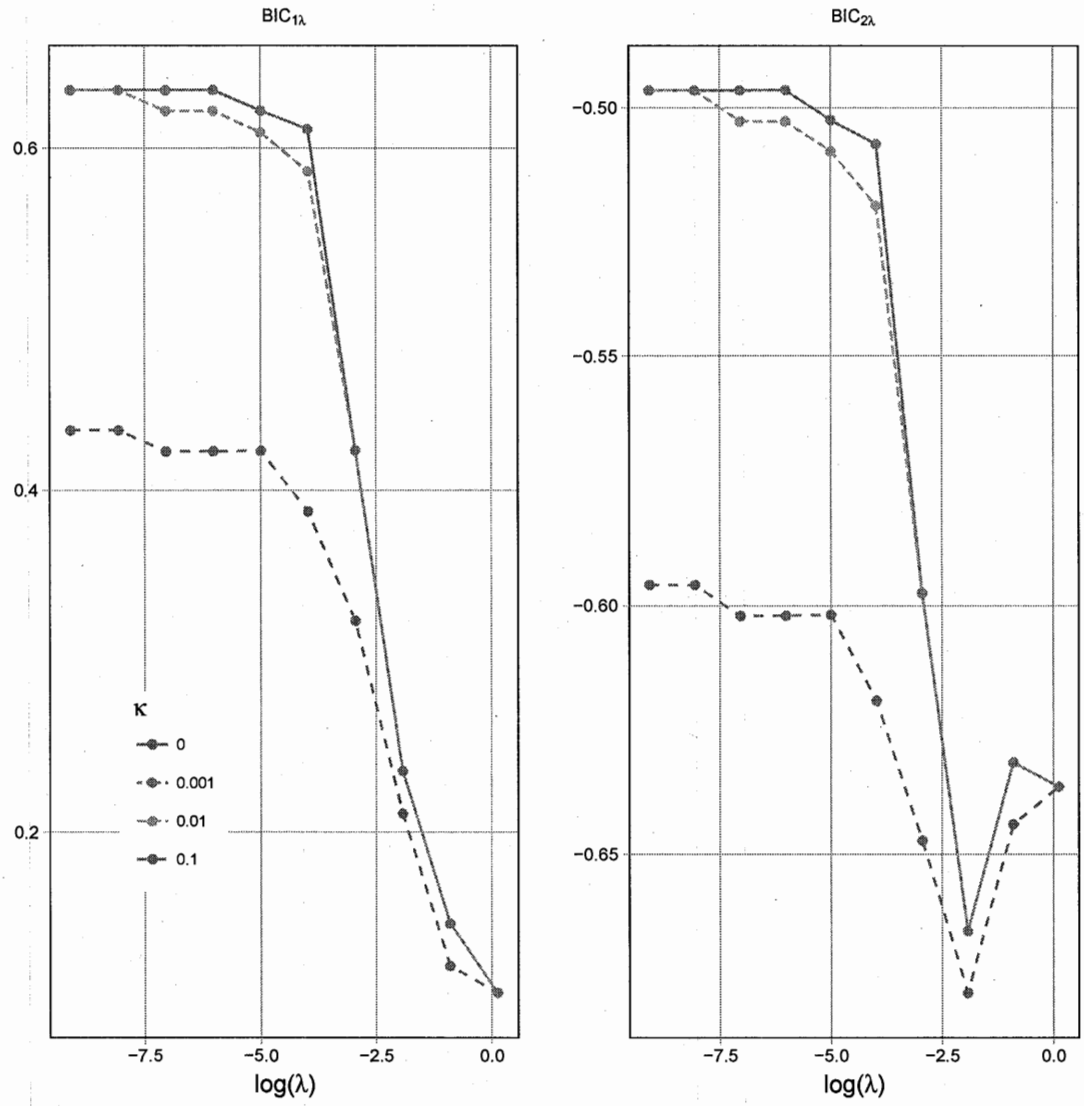

Figure 4.2: The values of $BIC_{1\lambda}$ and $BIC_{2\lambda}$, criteria as a function of $\log(\lambda)$ according to the different values of $\kappa \in (0, 0.001, 0.01, 0.1)$ for the sparse illustrative model

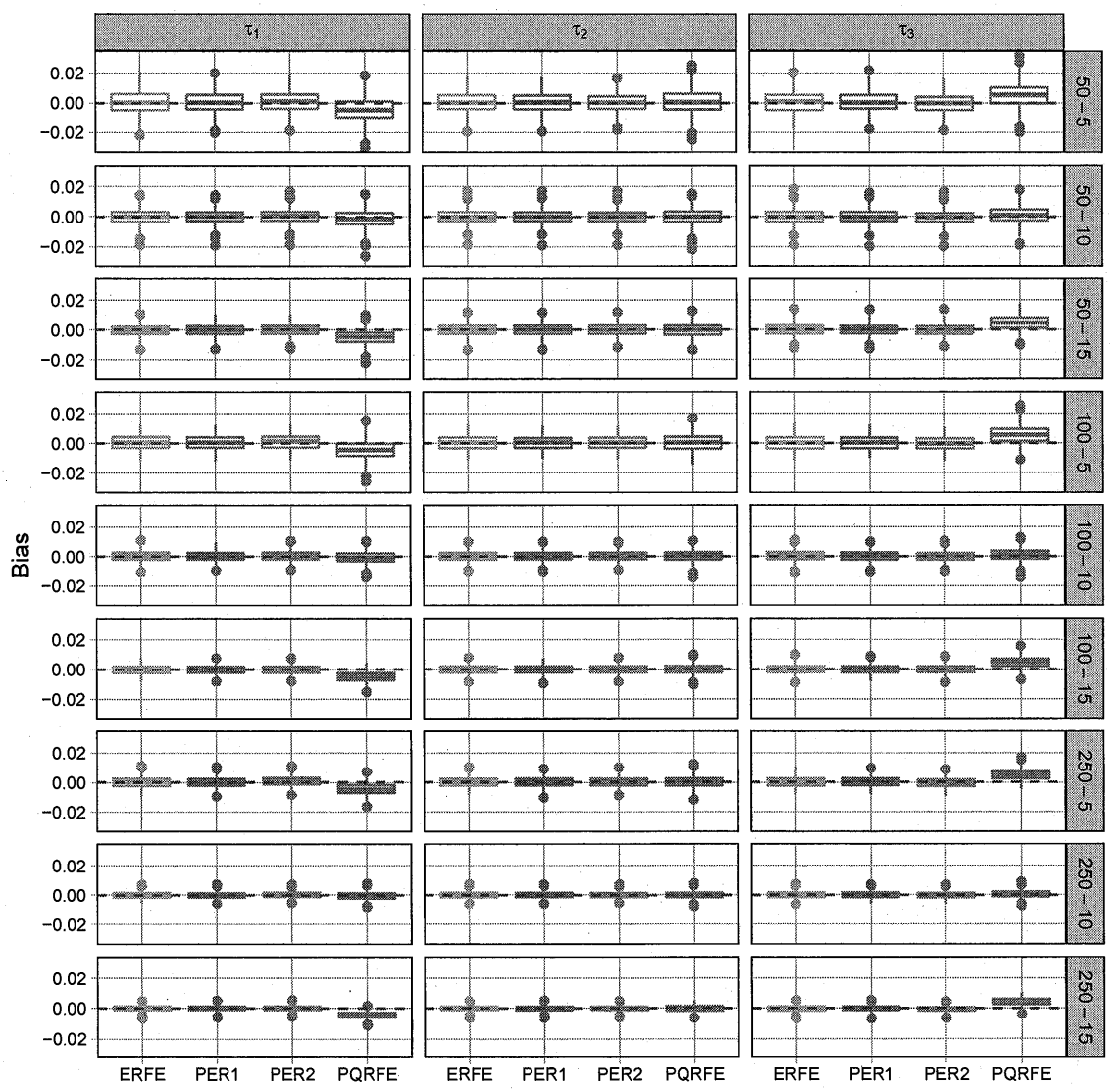

Figure 4.3: **Location-shift scenario** – Bias distribution (box-plot) of the estimators, ERFE, PER1, PER2, PQRFE, according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim \mathcal{N}(0, 1)$.

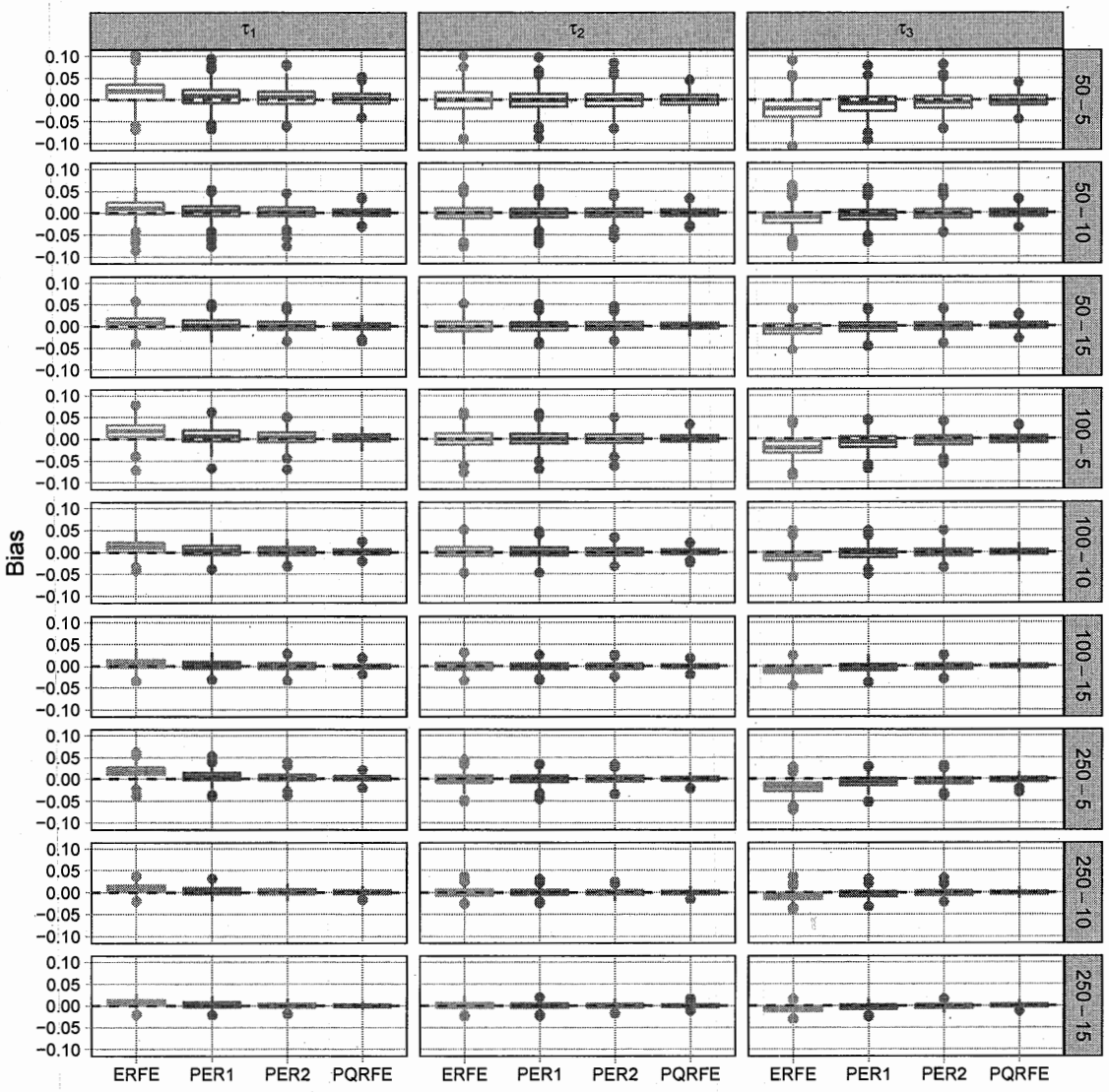

Figure 4.4: **Location-scale-shift scenario** – Bias distribution (box-plot) of the estimators, ERFE, PER1, PER2, PQRFE, according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim \mathcal{N}(0, 1)$.

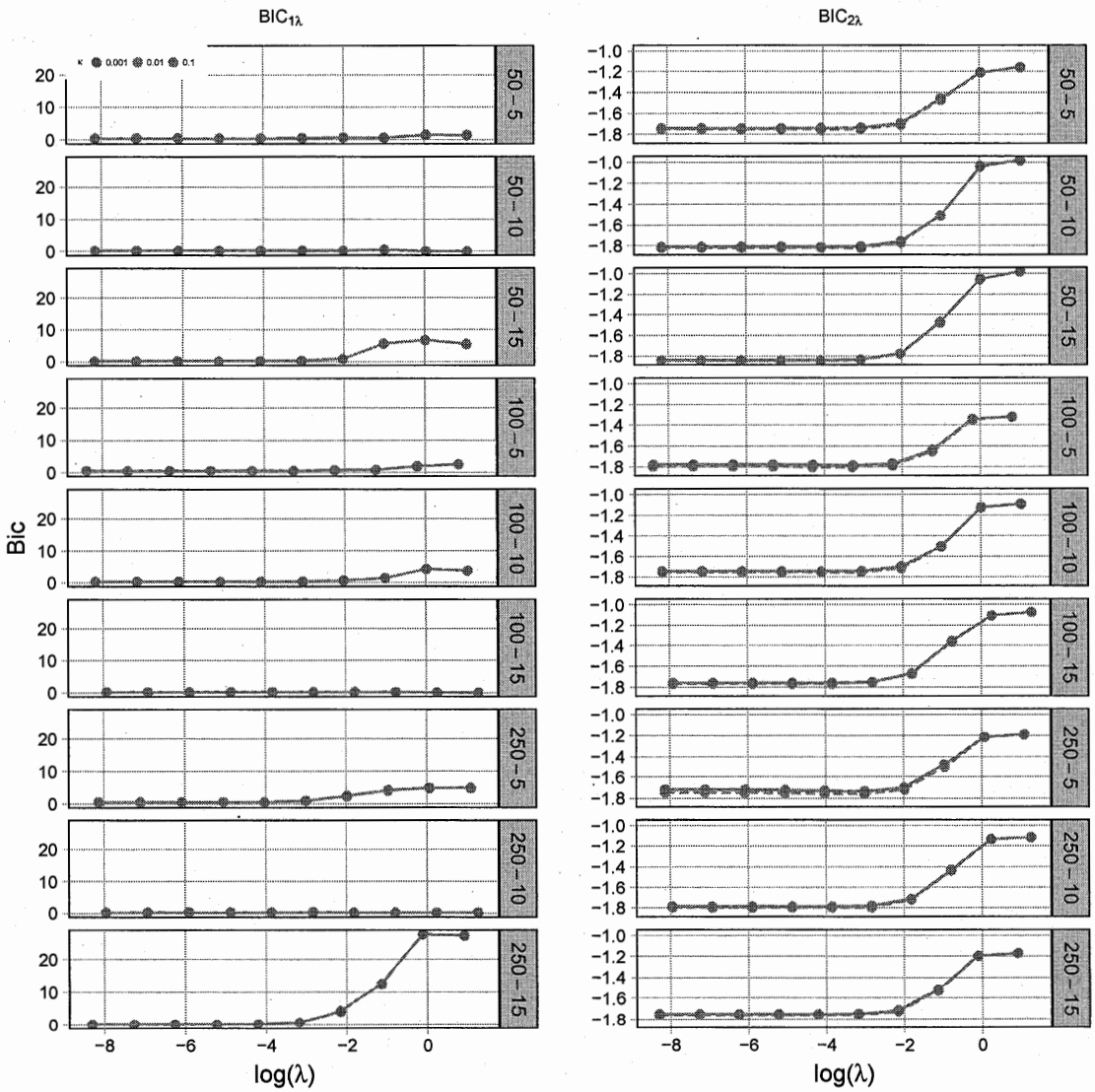

Figure 4.5: **Location-shift scenario** – BIC criteria plotted against the regularization parameter $\log(\lambda)$ according to the values of $\kappa = (0.001, 0.01, 0.1)$, for a single replication generated by a normal distribution, $\varepsilon \sim \mathcal{N}(0, 1)$.

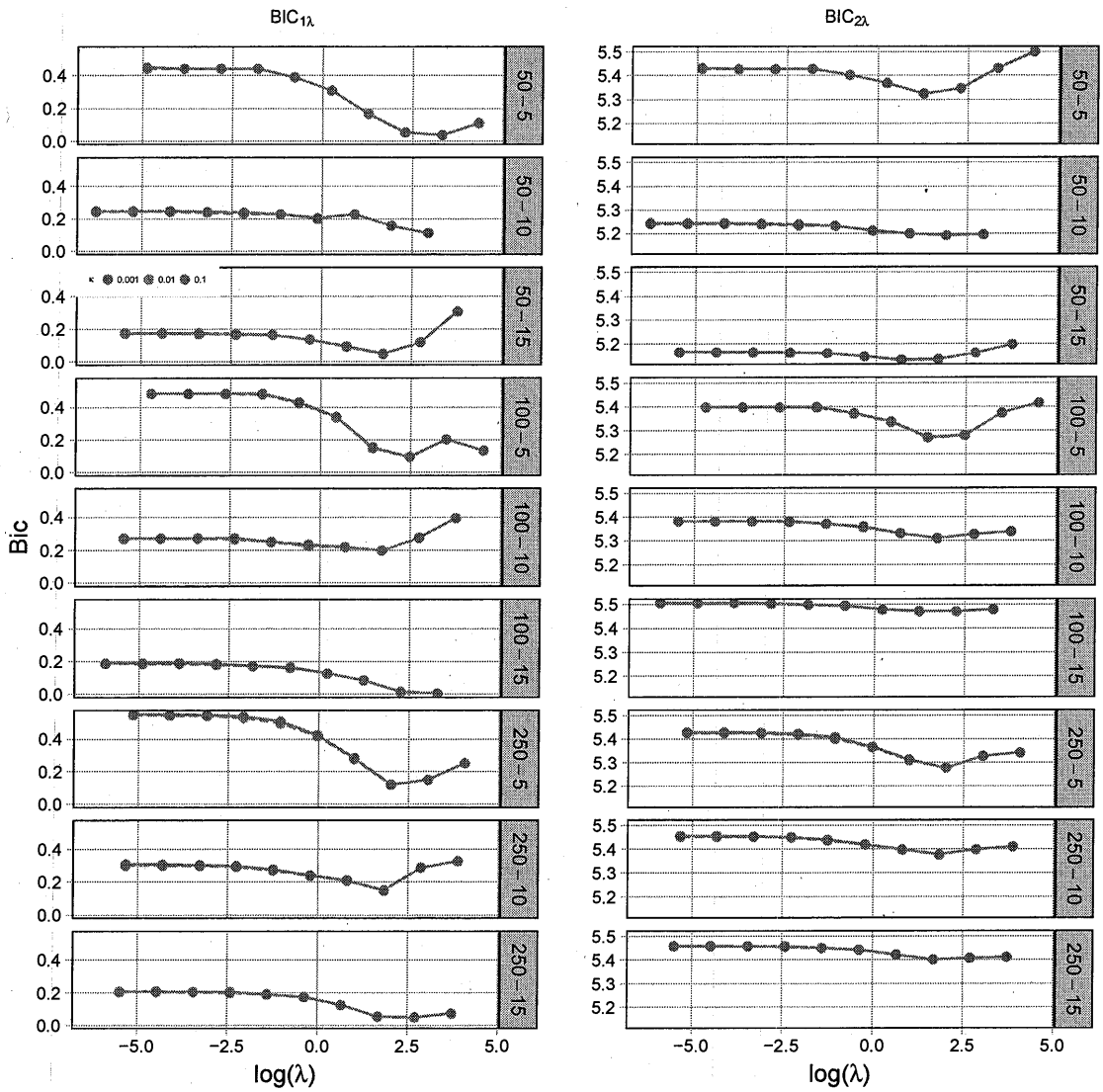

Figure 4.6: **Location-scale-shift scenario** – BIC criteria plotted against the regularization parameter $\log(\lambda)$ according to the values of $\kappa = (0.001, 0.01, 0.1)$, for a single replication generated by a normal distribution, $\varepsilon \sim \mathcal{N}(0, 1)$.

4.5 Application

Returns to schooling, also known as returns to education, is undoubtedly one of the most widely studied topic in empirical economics. An important component of the research on the subject is the investigation of the existence of individual-level heterogeneity in returns to education (Koop et Tobias, 2004). Research has suggested application of instrumental variable (IV) methods to account for this heterogeneity, (Cornwell et Rupert, 1988; Baltagi et Khanti-Akom, 1990). The difficulty of finding valid instruments and the question of verifying their underlying hypothesis hampers the application of the IV method (Greene, 2012; Baltagi, 2008). The fixed-effects model is by nature an IV method, therefore by extension our PERFE model. Moreover, the PERFE model allows inference of time-invariant covariates. In this section, we apply the PERFE model to study the individual-level heterogeneity in returns to schooling.

We replicated Baltagi and Khanti-Akon's study (Baltagi et Khanti-Akom, 1990) using the Panel Study of Income Dynamics (PSID) dataset (Cornwell et Rupert, 1988). The dataset is a cohort of 595 individuals observed over the period 1976–1982. The respondents, aged between 18 and 65 in 1976, are those who reported a positive wage in some private, non-farm employment for all 7 years, (Cornwell et Rupert, 1988). We also consider the PQRFE model in our analysis.

The log wage is the response variable and is regressed on years of education (ED), weeks worked (WKS), years of full-time work experience (EXP), occupation (OCC=1, if the individual is in a blue-collar occupation), residence (SOUTH = 1, SMSA = 1, if the individual resides in the South, or in a standard metropolitan statistical

area), marital status (MS = 1, if the individual is married), industry (IND = 1, if the individual works in a manufacturing industry), sex and race (FEM=1, BLK=1, if individual is female or black) and union coverage (UNION = 1, if the individual's wage is set by a union contract), (Baltagi, 2008). The corresponding Mincer equation (Lemieux, 2006) for both methods PERFE and PQRFE is

$$\begin{aligned}\mu_\tau(\log(\text{Wage}_{ij})) \text{ or } q_\tau(\log(\text{Wage}_{ij})) = \beta_{0\tau} &+ \beta_1 \text{ED}_{ij} + \beta_2 \text{FEM}_{ij} + \beta_3 \text{BLK}_{ij} \\ &+ \beta_4 \text{WKS}_{ij} + \beta_5 \text{EXP}_{ij} + \beta_6 \text{EXP}_{ij}^2 + \beta_7 \text{OCC}_{ij} \\ &+ \beta_8 \text{IND}_{ij} + \beta_9 \text{IND}_{ij} + \beta_{10} \text{SOUTH}_{ij} \\ &+ \beta_{11} \text{SMSA}_{ij} + \beta_{12} \text{MS}_{ij} + \alpha_i.\end{aligned}$$

We compute the solution for a fine grid with K = 100 different values of λ. Then, we estimated the model for all λ's in the grid and selected the optimal λ using $BIC_{2\lambda}$ criterion with $\kappa = 0.01$. As the regularization parameter $\lambda = \sigma_\varepsilon/\sigma_\alpha$ suggested by Koenker (2004) is not available for real dataset application, we used a similar grid to estimate the PQRFE model and select its optimal regularization parameter based on $BIC_{2\lambda}$ criterion. **Figure 4.10** plots $BIC_{2\lambda}$ as a function of $\log(\lambda)$. We observe that in both models, PERFE and PQRFE, the best regularization parameter corresponds to the model with fixed-effects.

We did not compute $BIC_{1\lambda}$ criterion because its expression is a function of the ERFE estimator (Barry *et al.*, 2018b) that does not estimate time-invariant covariates. Moreover, the PERFE does not estimate asymptotic standard error of time-invariant covariates. Instead, we generated the confidence intervals of the PERFE model using bootstrap. We generated 500 samples (same size as the original dataset) with replacement and estimated the PERFE model on each sample. Then, the confidence intervals is computed using the percentiles $(q_{0.025}, q_{0.975})$ of the bootstrap

distribution of the PERFE estimator. The confidence intervals of the PQRFE model is also generated by bootstrap. We estimated both models using 91 asymmetric points $(0.05, 0.06, 0.07, \ldots, 0.95)$. The results are presented in **Figure 4.7-4.9** and **Table 4.1**.

Notice that the point estimate of the PERFE model and that of the PQRFE model are not comparable. For a given τ, both methods do not estimate the same percentile (expectile or quantile). In view of the above, analysis and interpretation of PERFE and PQRFE results must be done globally, by looking at the global scale and trend of the heterogeneity of the effects according to the level τ-axis.

Figure 4.7 shows the heterogeneity of the effect of time-invariant covariates on wage distribution. Apart from education (ED), the effect of the covariates are relatively similar, having the same scale and similar pattern, for both models, PERFE and PQRFE. For example, the results of the PERFE model show a significant effect of ethnicity (black vs others) on wage. The effect is relativly on the same rate on the left, right and center of the wage distribution. That is, black with low, middle or high wage are disadvantaged in the same way. Although it is difficult to determine a trend from the PQRFE model, we observe a similar trend. The ethnicity effect on the wage distribution is significant and relatively uniform on the whole distribution. **Figure 4.8** and **Figure 4.9** report the heterogeneity of the effect of the other covariates on wage structure.

In addition to looking at the size and direction of the effect, it is important to identify the effects that are significant. To see more closely the effects, their sign, magnitude and significance, we extracted the results of some asymmetric points $\tau \in (0.1, 0.25, 0.5, 0.75, 0.9)$. Results of some important covariates are presented in

Table 4.1 and the remaining are in the **Supplementary material I**. The results from that table allow us to see from another perspective the utility of both methods in data analysis. It can be seen that the effect of some covariates is not significant at any point of the distribution, while others have a significant effect on the left tail but not on the right tail of the distribution, while some covariates have a significant effect on the whole distribution of the response variable. This is the case, for example, with education. Education has a significant effect between 10% and 13%, depending on whether one has low/high wages in the PERFE model and a significant effect between 5% to 6% depending on the wage level for the PQRFE model. While working in a manufacturing industry (IND) has no significant effect on wage change. On the other hand, according to the results of the PERFE model, the effect of being unionized is significant only to the left tail of the wage distribution. Being part of a union (UNION) is more advantageous for low wages than for high wages. This finding, on the heterogeneity of the effect of unions on the structure of wages, is consistent with the results obtained by Card (1996). Finally, both methods report a significant effect of the OCC covariate on the entire distribution of the wage. This difference between blue and white-collar is more pronounced in the group with low wage, it is -7.6% when $\tau = 0.1$ and -3.7% $\tau = 0.9$ in the PERFE model. We noticed that the results of the PQRFE model are similar, if not identical, to the results of the classical QR model. Both models the PQRFE and QR display same scale and same pattern. The results of the QR model can be found in (Barry *et al.*, 2018b).
Analysis of the results showed existence of a subject-level unobserved heterogeneity and heterogeneous covariates effects in returns to schooling. The PERFE model goes beyond the average or the median to capture the dynamics of the effects on the structure of the wage distribution. In addition, the PERFE model accounts for the subject-level unobserved heterogeneity and allows time-invariant inference. Given

its attractive properties, future research should investigate the application of the PERFE model to other penalties, such as elastic net penalty (Hui et Trevor, 2005).

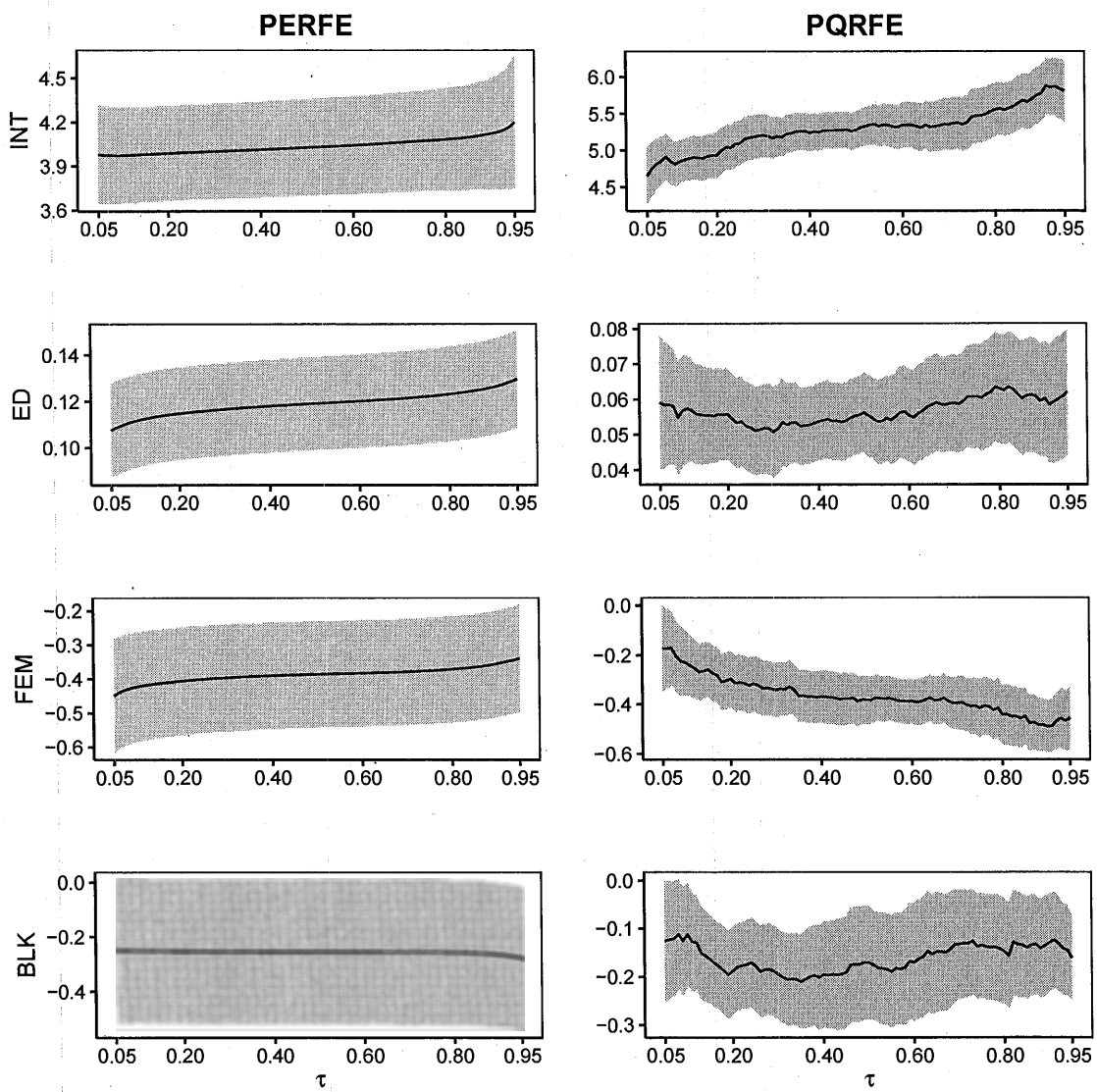

Figure 4.7: PERFE and PQRFE estimated coefficients with estimated confidence intervals, Part I.

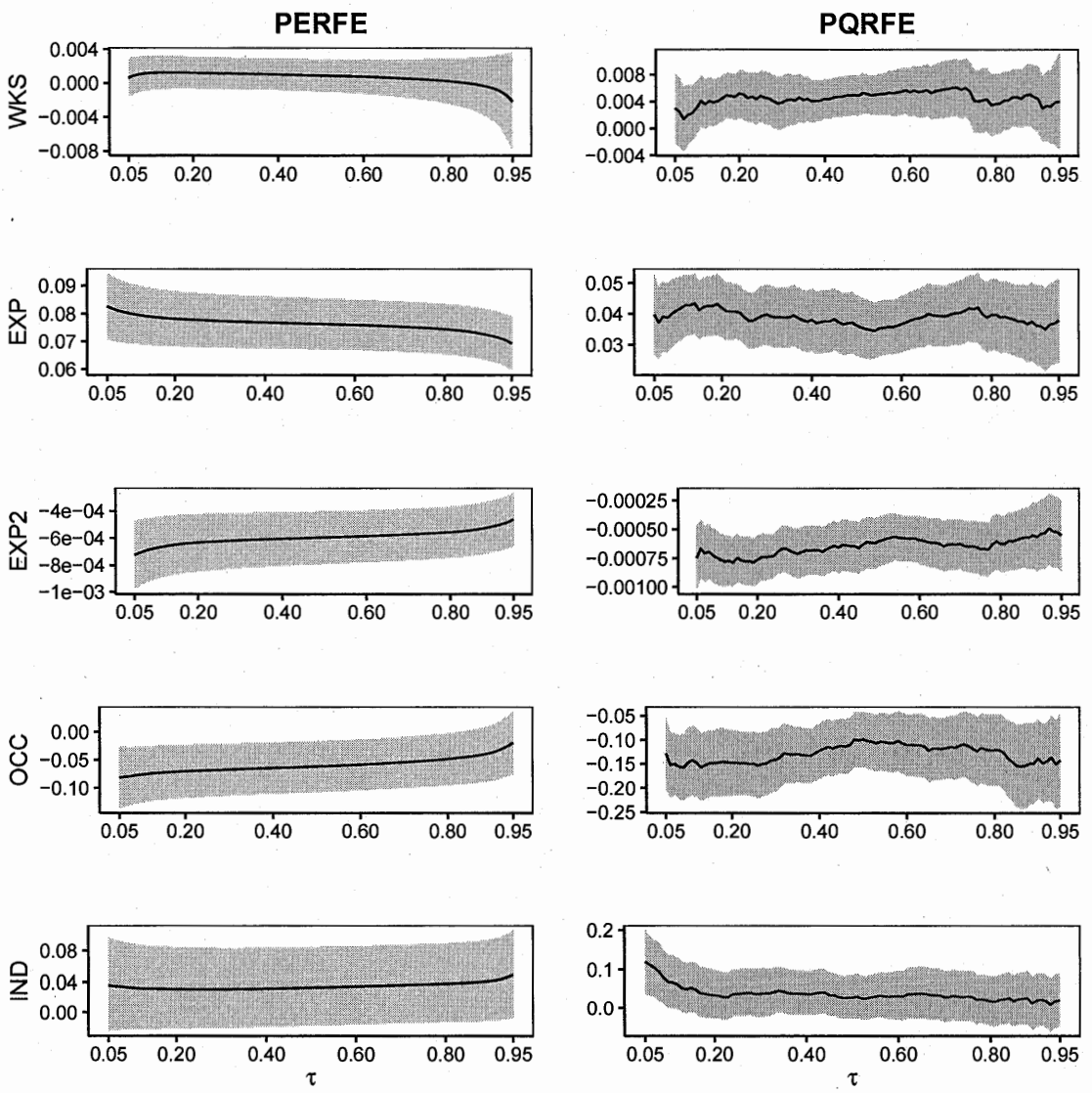

Figure 4.8: PERFE and PQRFE estimated coefficients with estimated confidence intervals, Part II.

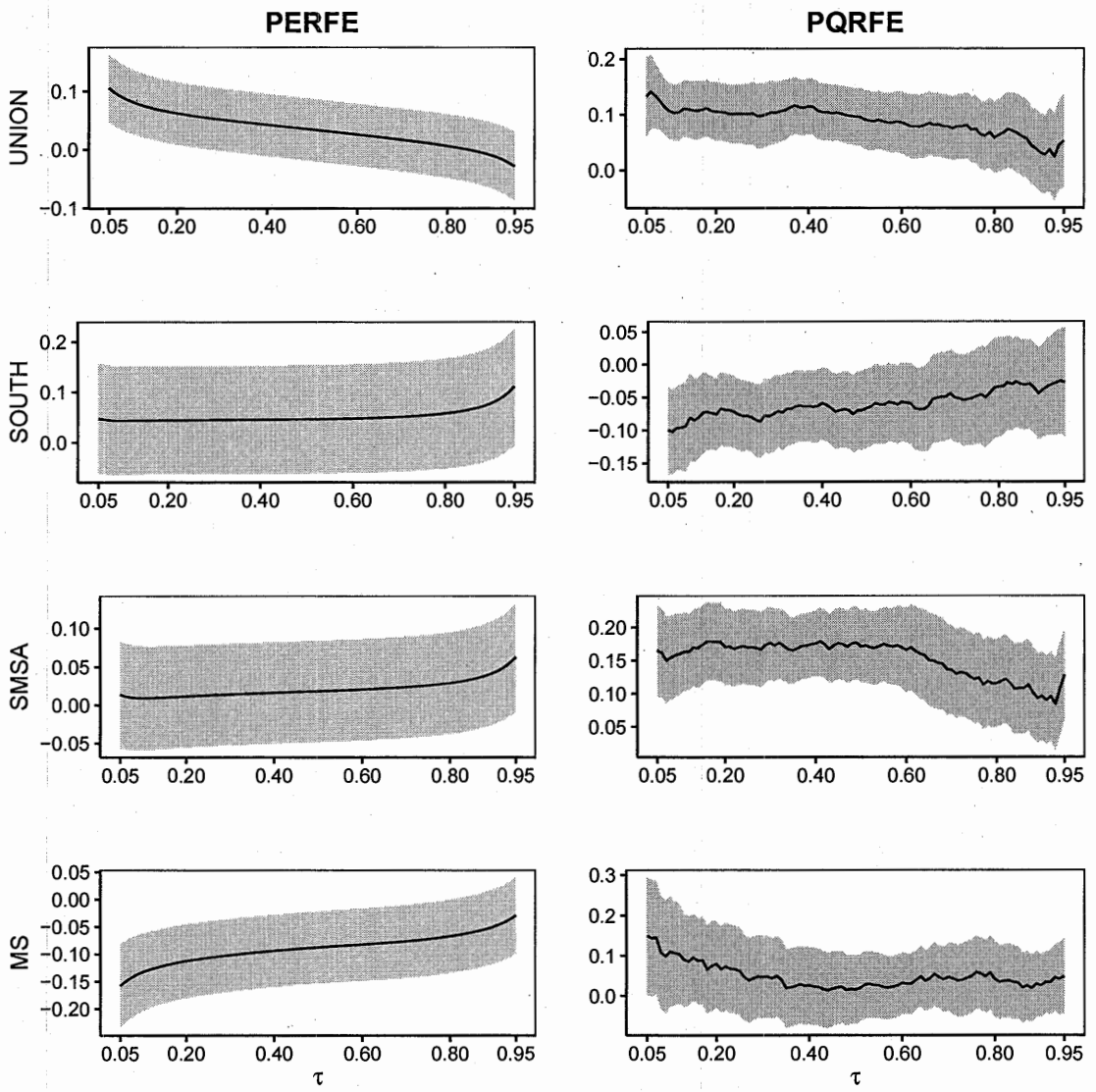

Figure 4.9: PERFE and PQRFE estimated coefficients with estimated confidence intervals, Part III.

Tableau 4.1: Parameters estimates (Est) and p-values obtained from the PERFE and PQRFE methods at five percentiles, $\tau = (0.1, 0.25, 0.5, 0.75, 0.9)$.

Var	τ	0.1	0.25	0.5	0.75	0.9
ED	PERFE	0.1114***	0.1158***	0.1190***	0.1221***	0.1262***
	PQRFE	0.0600***	0.0500***	0.0600***	0.0600***	0.0600***
IND	PERFE	0.0333	0.0312	0.0335	0.0372	0.0415
	PQRFE	0.0700*	0.0400	0.0300	0.0300	0.0200
OCC	PERFE	-0.0759***	-0.0686***	-0.0618***	-0.0519**	-0.0372
	PQRFE	-0.1500***	-0.1500***	-0.1000***	-0.1200***	-0.1400***
UNION	PERFE	0.0827***	0.0564**	0.0338	0.0117	-0.0104
	PQRFE	0.1100***	0.1000***	0.1000***	0.0700***	0.0300

***significant at 1% level, **significant at 5% level, *significant at 10% level.

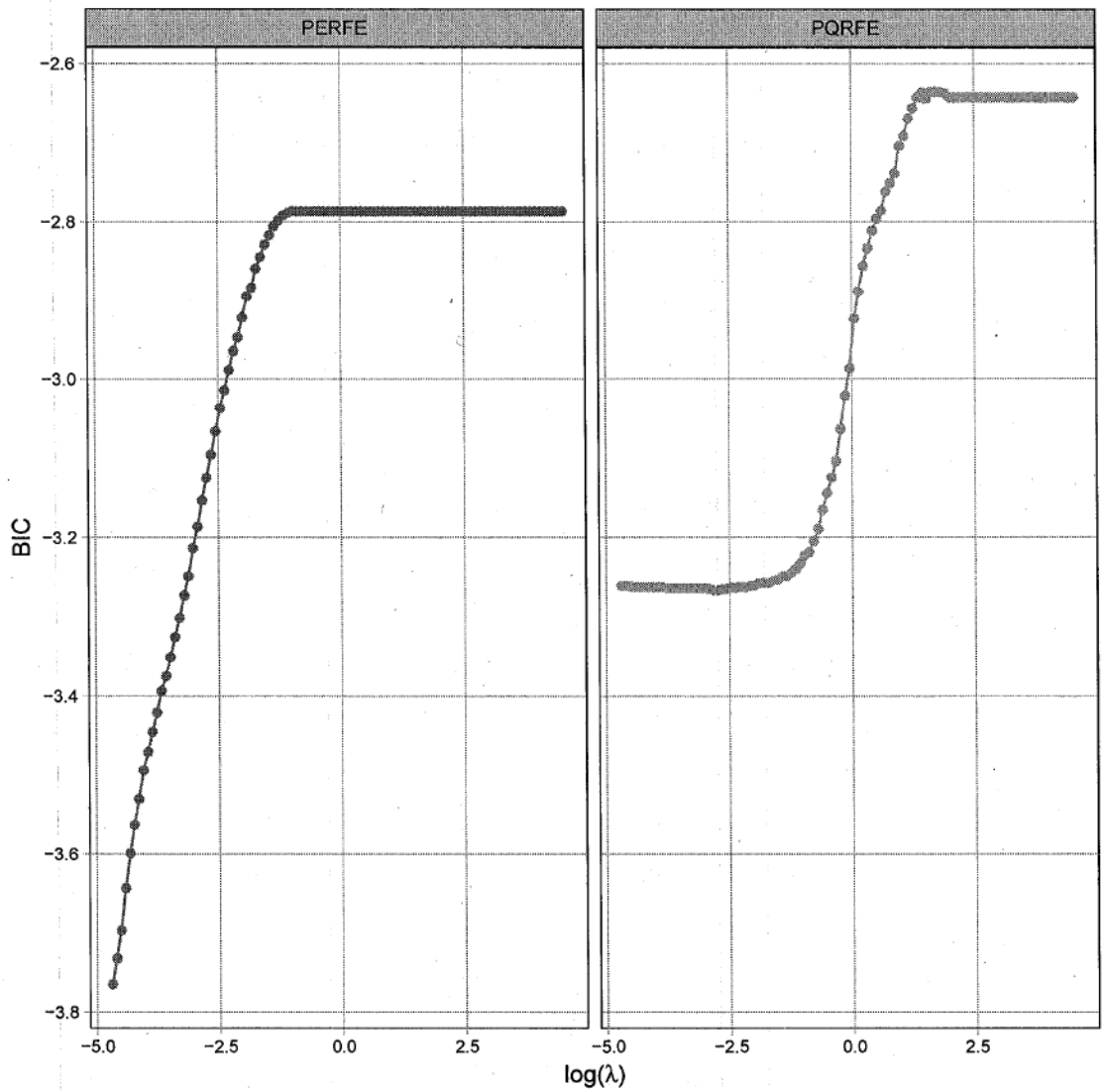

Figure 4.10: Values of the BIC criterion $BIC_{2\lambda}$ plotted against $\log(\lambda)$ according to the PERFE and PQRFE models.

4.6 Conclusion

In this paper, we introduced the penalized weighted asymmetric least squares regression for longitudinal data with fixed-effects. The PERFE model uses the l_1 penalty to control the variability introduced by the incidental parameter and at the same time prevents overfitting. We resort to special treatment to compute the solution path of the regularization parameter and rely on the Bayesian information criterion (BIC) to select the optimal regularization parameter.

The simulation results show that the PERFE estimator has lower bias and better efficiency and outperforms its competitors (ERFE, PQRFE), particularly in the location-shift scenario. We fit the PERFE model to the Panel Study of Income Dynamics to study the returns to schooling. The results reveal the presence of an individual-level heterogeneity in returns to schooling.

Today, high-dimensional longitudinal dataset is available in several consortium such as the UK10K consortium. The penalized ERFE model can be fit to such data to detect important covariates for the conditional mean and important covariates for the conditional scale/variance, using techniques from SALES and COSALES (Gu et Zou, 2016). Future research should also investigate application of other penalties such as Elastic net, SCAD and MCP to the ERFE model.

4.7 Appendix

4.7.1 Supplementary Material I : Additional results

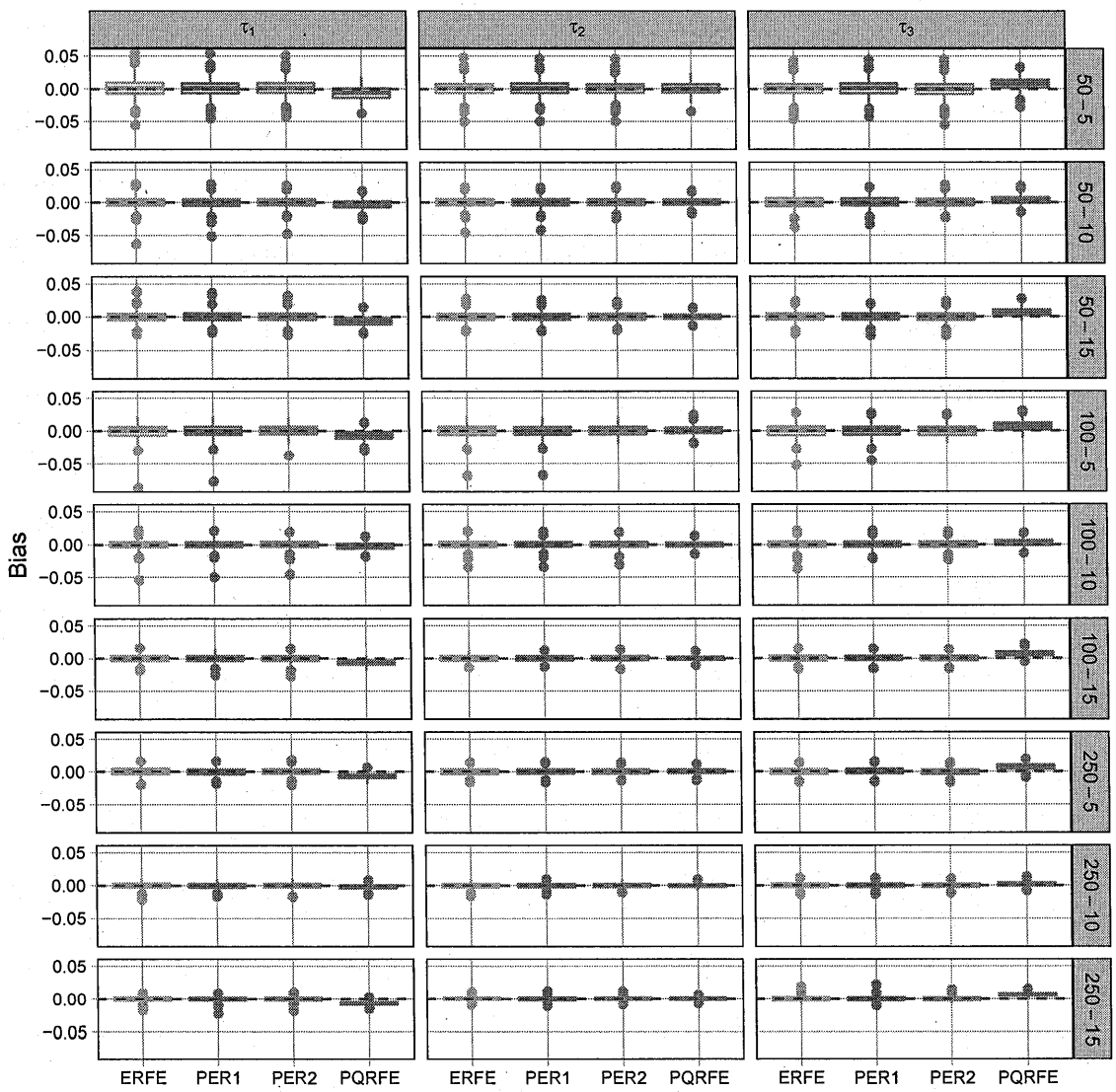

Figure 4.11: **Location-shift scenario** – Bias distribution (box-plot) of the estimators, ERFE, PER1, PER2, PQRFE, according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim t_3$.

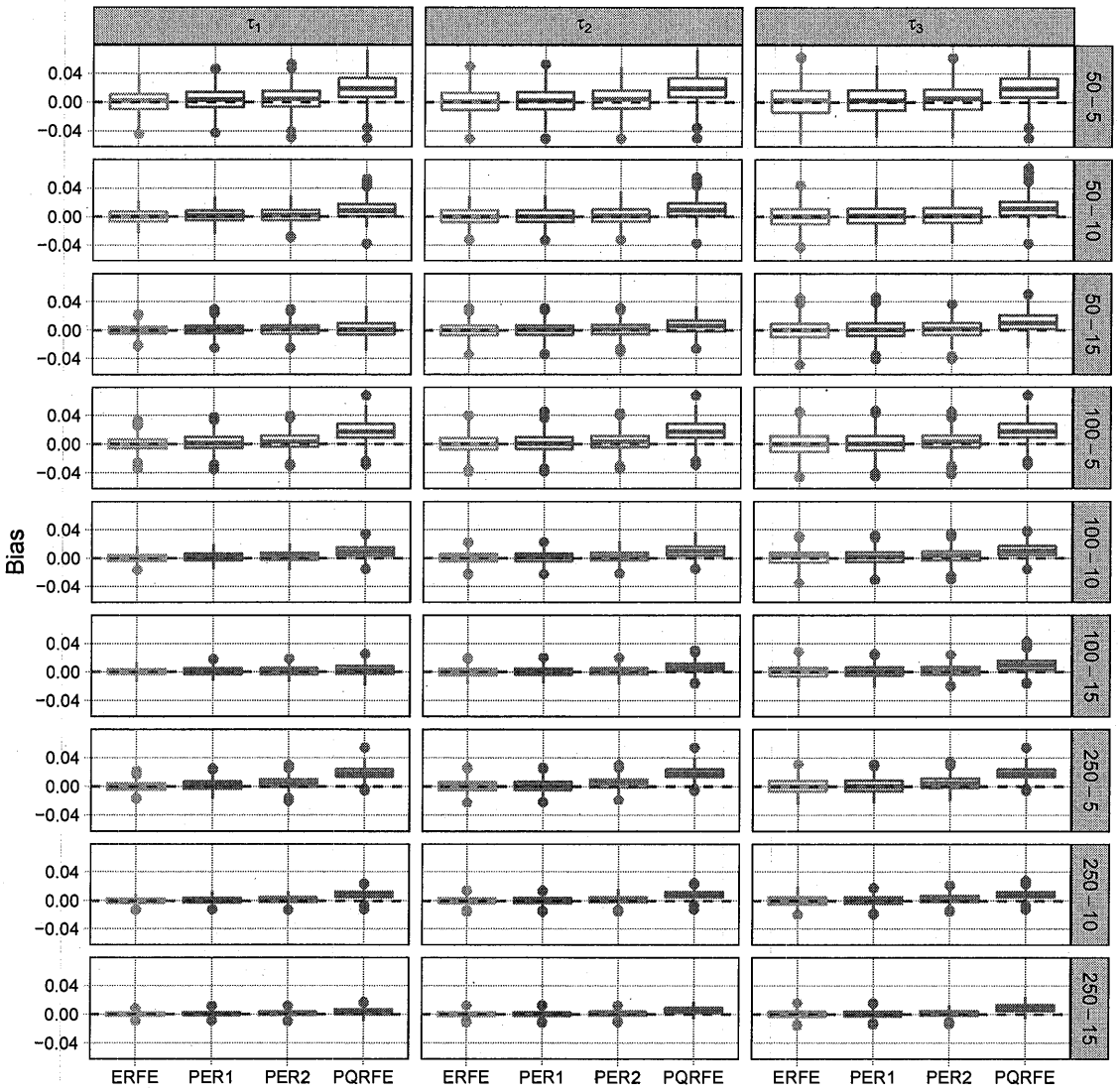

Figure 4.12: **Location-shift scenario** – Bias distribution (box-plot) of the estimators, ERFE, PER1, PER2, PQRFE, according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim \chi^3_2$.

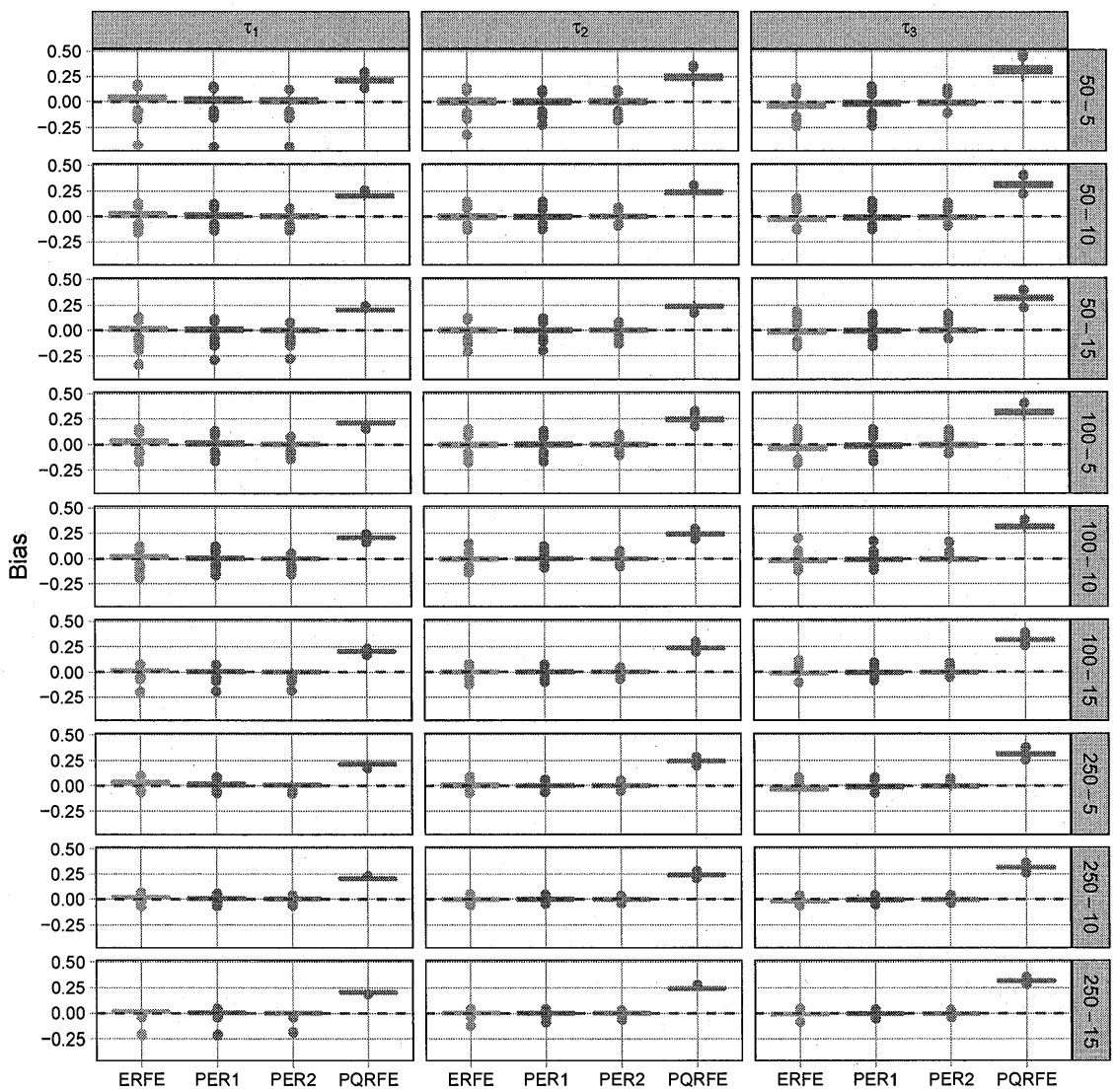

Figure 4.13: **Location-scale-shift scenario** – Bias distribution (box-plot) of the estimators, ERFE, PER1, PER2, PQRFE, according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim t_3$.

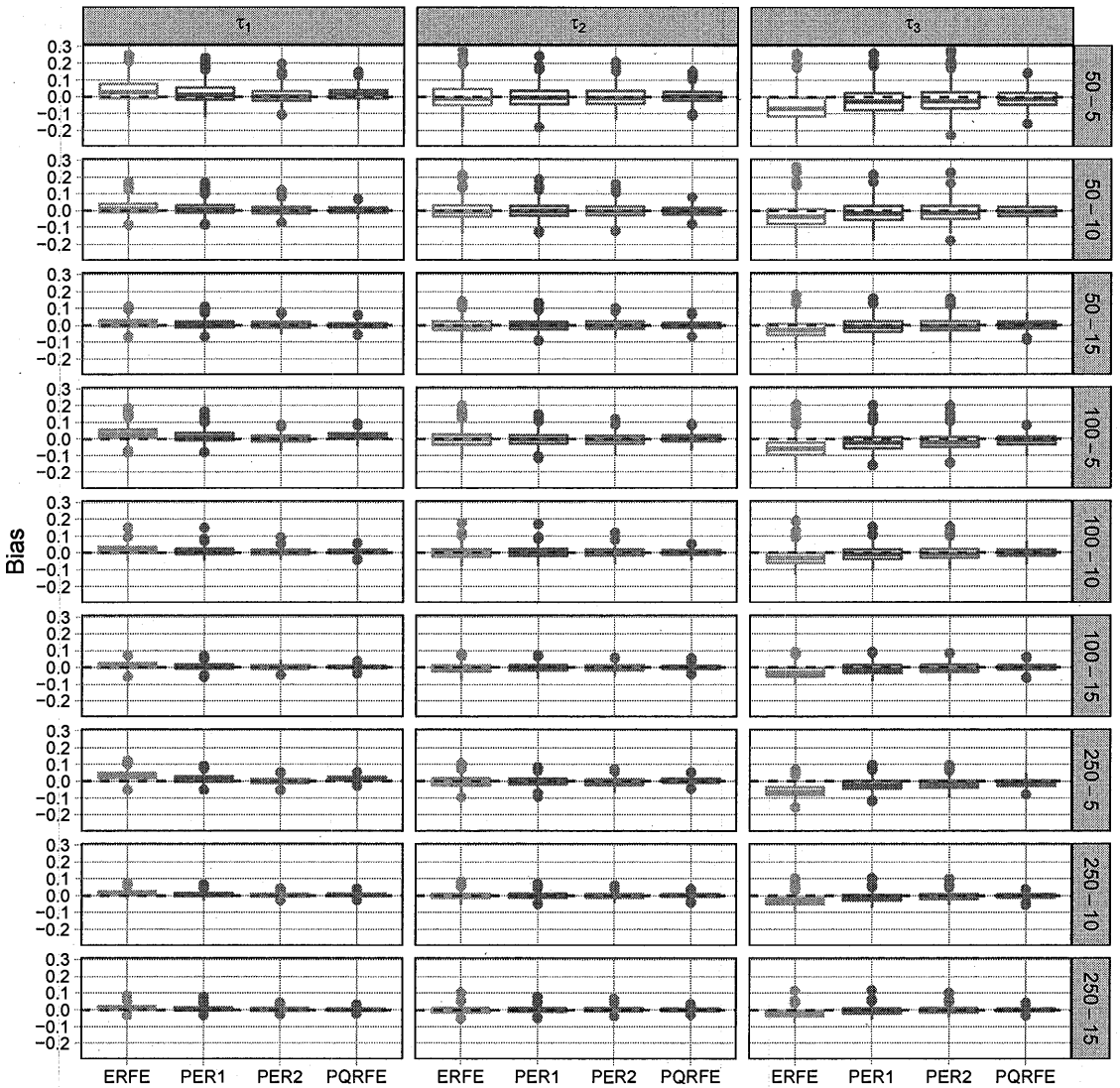

Figure 4.14: **Location-scale-shift scenario** – Bias distribution (box-plot) of the estimators, ERFE, PER1, PER2, PQRFE, according to $\tau = (\tau_1, \tau_2, \tau_3)$, the number of subjects, $n = (250, 100, 50)$, the number of within-subject observations, $m = (5, 10, 15)$, and the error term, $\varepsilon \sim \chi^3_2$.

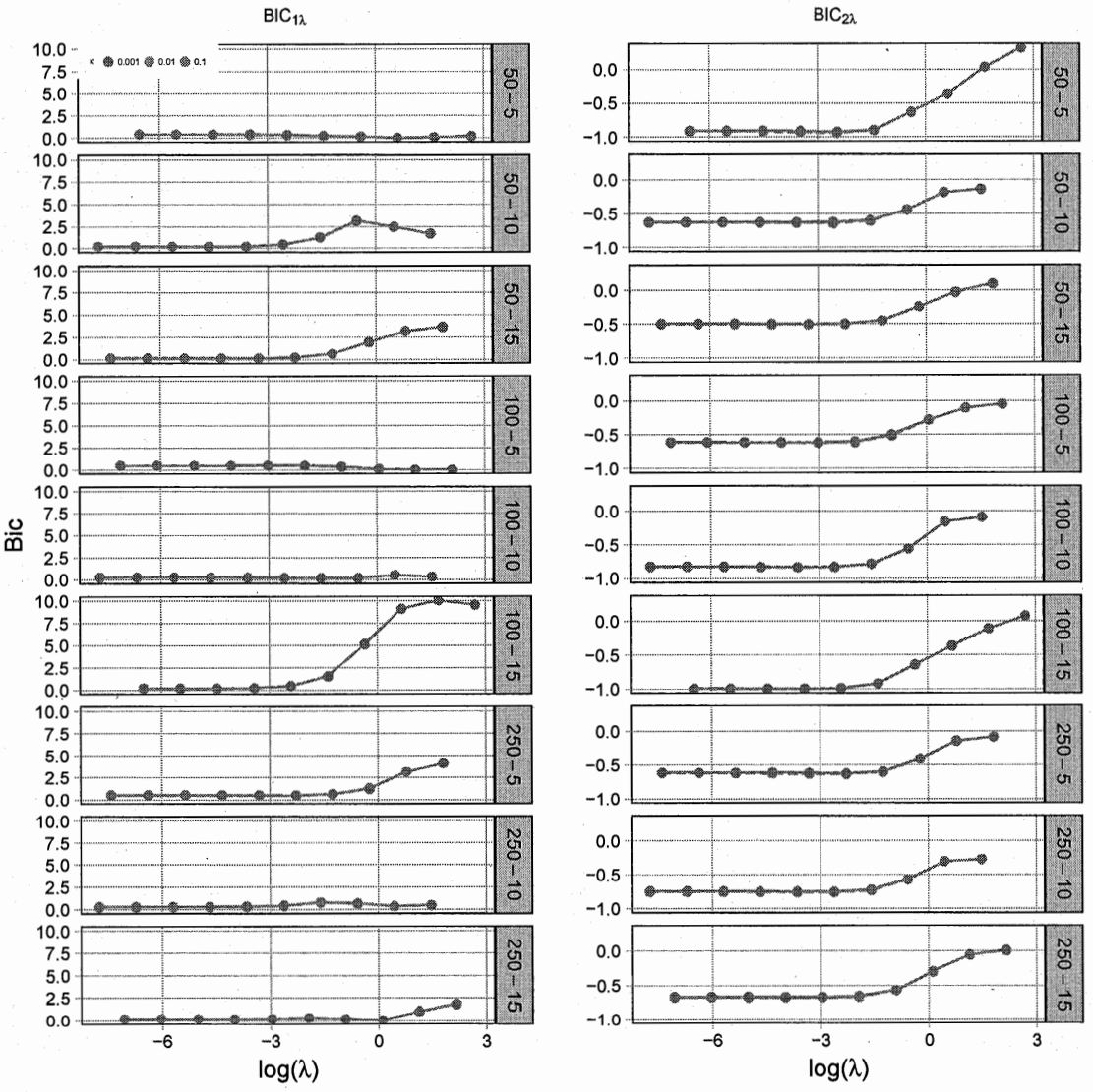

Figure 4.15: **Location-shift scenario** – BIC criteria plotted against the regularization parameter $\log(\lambda)$ according to the values of $\kappa = (0.001, 0.01, 0.1)$, for a single replication generated by a Student distribution, $\varepsilon \sim t_3$.

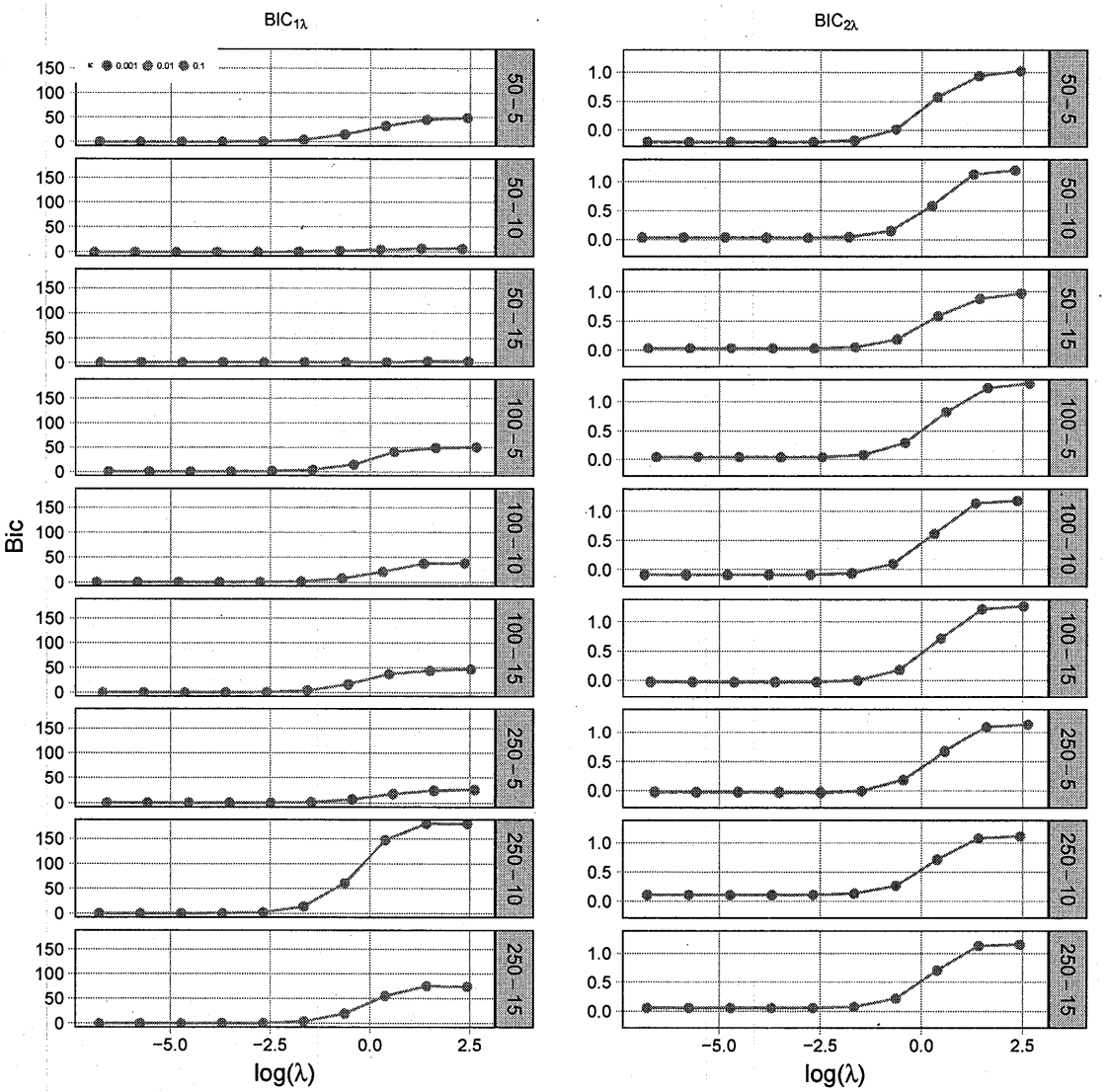

Figure 4.16: **Location-shift scenario** – BIC criteria plotted against the regularization parameter $\log(\lambda)$ according to the values of $\kappa = (0.001, 0.01, 0.1)$, for a single replication generated by a Chi-Square distribution, $\varepsilon \sim \chi^3_2$.

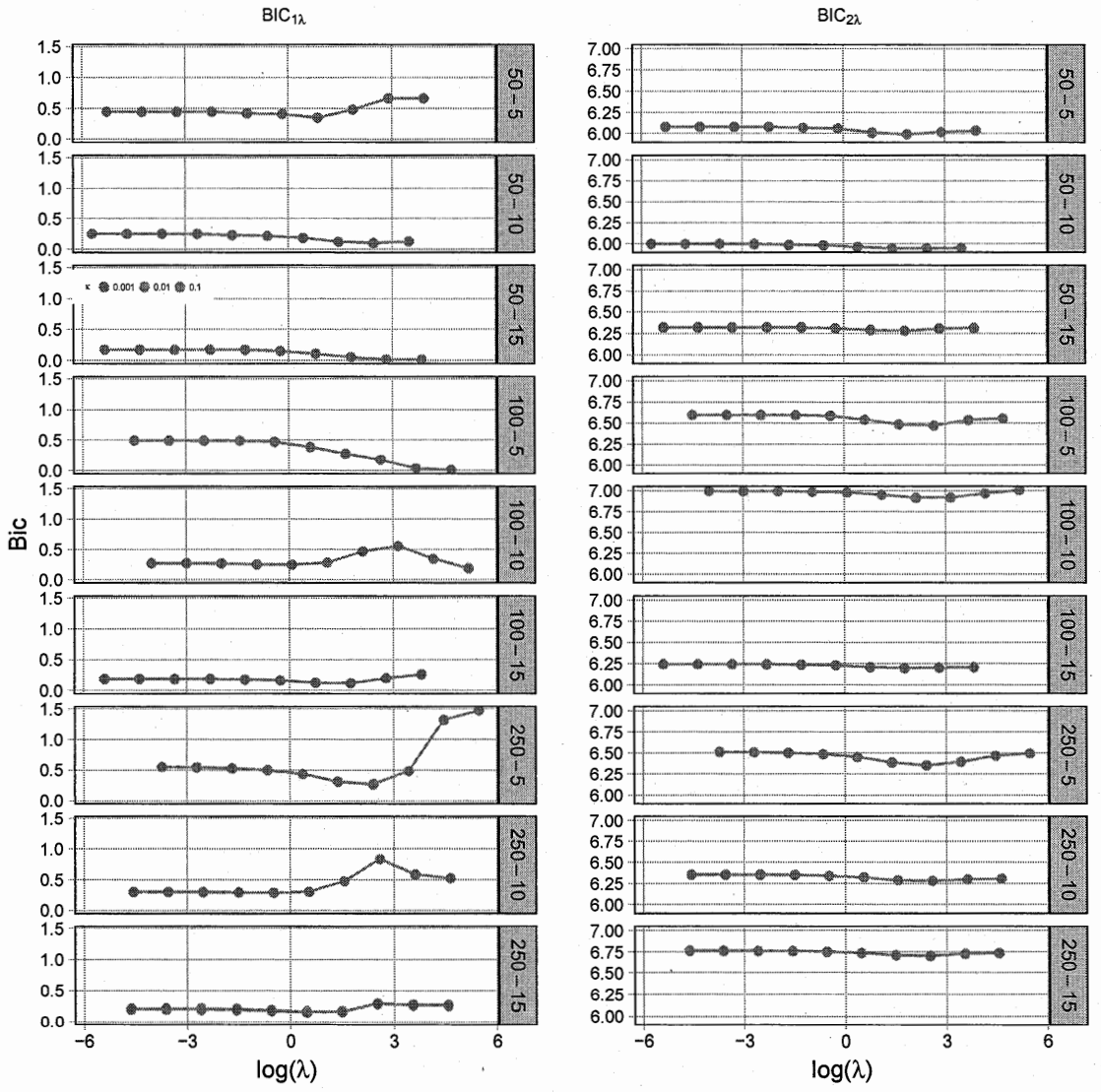

Figure 4.17: **Location-scale-shift scenario** – BIC criteria plotted against the regularization parameter $\log(\lambda)$ according to the values of $\kappa = (0.001, 0.01, 0.1)$, for a single replication generated by a Student distribution, $\varepsilon \sim t_3$.

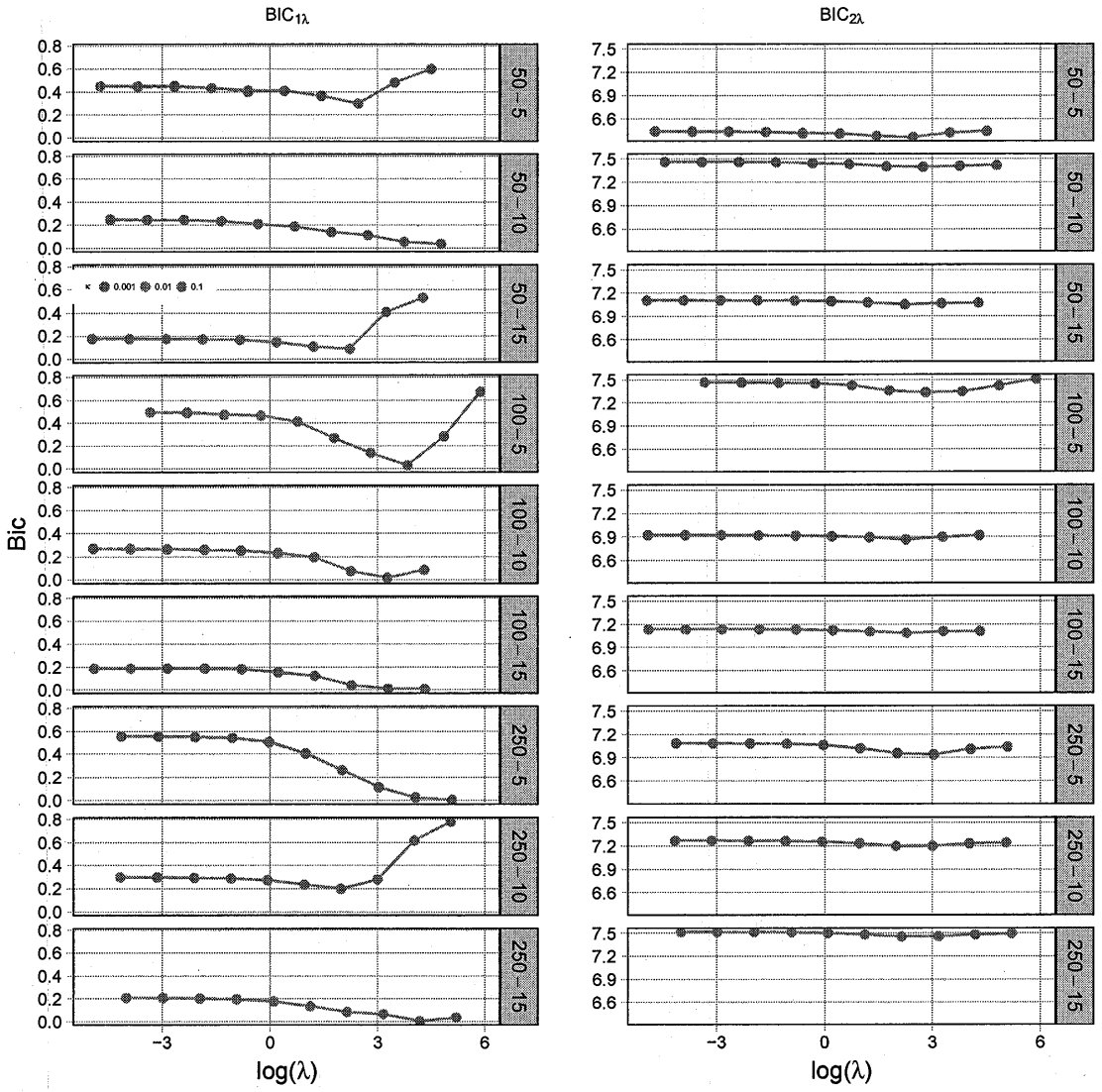

Figure 4.18: **Location-scale-shift scenario** – BIC criteria plotted against the regularization parameter $\log(\lambda)$ according to the values of $\kappa = (0.001, 0.01, 0.1)$, for a single replication generated by a Chi-Square distribution, $\varepsilon \sim \chi^3_2$.

Tableau 4.2: Parameters estimates (Est) with their 95% confidence intervals (CI) and p-values obtained from the PERFE and PQRFE methods at five percentiles, $\tau = (0.1, 0.25, 0.5, 0.75, 0.9)$.

	τ	0.1			0.25			0.5			0.75			0.9		
Var	Method	Est	CI	P	Est	CI	P	Est	CI	P	Est	CI	P	Est	CI	P
Int	PERFE	3.97	(3.65, 4.30)	0.00	4.00	(3.67, 4.32)	0.00	4.03	(3.70, 4.35)	0.00	4.07	(3.73, 4.41)	0.00	4.12	(3.75, 4.50)	0.00
	PQRFE	4.85	(4.54, 5.15)	0.00	5.10	(4.79, 5.40)	0.00	5.28	(5.03, 5.53)	0.00	5.46	(5.09, 5.83)	0.00	5.77	(5.39, 6.15)	0.00
BLK	PERFE	-0.25	(−0.51, 0.01)	0.05	-0.26	(−0.52, 0.00)	0.05	-0.26	(−0.52, 0.00)	0.05	-0.26	(−0.52, 0.00)	0.05	-0.27	(−0.53, −0.01)	0.04
	PQRFE	-0.11	(−0.22, −0.01)	0.03	-0.18	(−0.27, −0.08)	0.00	-0.17	(−0.29, −0.05)	0.00	-0.13	(−0.24, −0.03)	0.01	-0.13	(−0.23, −0.03)	0.01
ED	PERFE	0.11	(0.09, 0.13)	0.00	0.12	(0.10, 0.14)	0.00	0.12	(0.10, 0.14)	0.00	0.12	(0.10, 0.14)	0.00	0.13	(0.11, 0.15)	0.00
	PQRFE	0.06	(0.04, 0.07)	0.00	0.05	(0.04, 0.06)	0.00	0.06	(0.05, 0.07)	0.00	0.06	(0.05, 0.08)	0.00	0.06	(0.04, 0.08)	0.00
EXP	PERFE	0.08	(0.07, 0.09)	0.00	0.08	(0.07, 0.09)	0.00	0.08	(0.07, 0.09)	0.00	0.08	(0.07, 0.08)	0.00	0.07	(0.06, 0.08)	0.00
	PQRFE	0.04	(0.03, 0.05)	0.00	0.04	(0.03, 0.05)	0.00	0.04	(0.03, 0.05)	0.00	0.04	(0.03, 0.05)	0.00	0.04	(0.02, 0.05)	0.00
EXP2	PERFE	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00
	PQRFE	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00	0.00	(0.00, 0.00)	0.00
FEM	PERFE	-0.42	(−0.58, −0.26)	0.00	-0.40	(−0.56, −0.24)	0.00	-0.39	(−0.54, −0.23)	0.00	-0.38	(−0.53, −0.22)	0.00	-0.36	(−0.51, −0.20)	0.00
	PQRFE	-0.23	(−0.36, −0.10)	0.00	-0.32	(−0.43, −0.20)	0.00	-0.38	(−0.48, −0.29)	0.00	-0.42	(−0.53, −0.30)	0.00	-0.49	(−0.59, −0.39)	0.00
IND	PERFE	0.03	(−0.02, 0.09)	0.23	0.03	(−0.02, 0.08)	0.23	0.03	(−0.02, 0.08)	0.19	0.04	(−0.01, 0.09)	0.14	0.04	(−0.01, 0.09)	0.12
	PQRFE	0.07	(0.00, 0.14)	0.04	0.04	(−0.02, 0.10)	0.20	0.03	(−0.03, 0.09)	0.32	0.03	(−0.03, 0.10)	0.33	0.02	(−0.05, 0.09)	0.57
MS	PERFE	-0.13	(−0.20, −0.06)	0.00	-0.11	(−0.17, −0.04)	0.00	-0.09	(−0.15, −0.02)	0.01	-0.07	(−0.14, −0.01)	0.03	-0.05	(−0.12, 0.02)	0.14
	PQRFE	0.11	(−0.03, 0.25)	0.11	0.06	(−0.05, 0.17)	0.25	0.02	(−0.07, 0.10)	0.72	0.05	(−0.04, 0.14)	0.27	0.03	(−0.05, 0.10)	0.49
OCC	PERFE	-0.08	(−0.13, −0.03)	0.00	-0.07	(−0.12, −0.02)	0.00	-0.06	(−0.11, −0.02)	0.01	-0.05	(−0.10, −0.01)	0.02	-0.04	(−0.09, 0.01)	0.13
	PQRFE	-0.15	(−0.22, −0.08)	0.00	-0.15	(−0.22, −0.08)	0.00	-0.10	(−0.16, −0.04)	0.00	-0.12	(−0.19, −0.05)	0.00	-0.14	(−0.22, −0.06)	0.00
SMSA	PERFE	0.01	(−0.06, 0.08)	0.80	0.01	(−0.05, 0.08)	0.70	0.02	(−0.05, 0.08)	0.59	0.03	(−0.04, 0.09)	0.44	0.04	(−0.02, 0.11)	0.21
	PQRFE	0.16	(0.10, 0.22)	0.00	0.17	(0.11, 0.22)	0.00	0.17	(0.12, 0.23)	0.00	0.12	(0.06, 0.19)	0.00	0.10	(0.03, 0.16)	0.00
SOUTH	PERFE	0.04	(−0.06, 0.15)	0.42	0.04	(−0.06, 0.15)	0.40	0.05	(−0.06, 0.15)	0.39	0.05	(−0.05, 0.16)	0.32	0.08	(−0.03, 0.19)	0.16
	PQRFE	-0.08	(−0.15, −0.01)	0.02	-0.08	(−0.14, −0.03)	0.00	-0.07	(−0.13, −0.01)	0.03	-0.05	(−0.12, 0.02)	0.18	-0.04	(−0.11, 0.03)	0.27
UNION	PERFE	0.08	(0.03, 0.14)	0.00	0.06	(0.00, 0.11)	0.03	0.03	(−0.02, 0.09)	0.20	0.01	(−0.04, 0.07)	0.66	-0.01	(−0.07, 0.05)	0.71
	PQRFE	0.11	(0.06, 0.16)	0.00	0.10	(0.05, 0.15)	0.00	0.10	(0.04, 0.15)	0.00	0.07	(0.01, 0.13)	0.01	0.03	(−0.04, 0.10)	0.38
WKS	PERFE	0.00	(0.00, 0.00)	0.23	0.00	(0.00, 0.00)	0.20	0.00	(0.00, 0.00)	0.34	0.00	(0.00, 0.00)	0.73	0.00	(0.00, 0.00)	0.77
	PQRFE	0.00	(0.00, 0.01)	0.11	0.00	(0.00, 0.01)	0.01	0.01	(0.00, 0.01)	0.00	0.00	(0.00, 0.01)	0.07	0.00	(0.00, 0.01)	0.07

4.7.2 Supplementary Material II : Proof of the Theorems

Proof of Theorem 4.3.1.

Let $\mu_{ij\tau} = \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\beta}_\tau + \boldsymbol{z}_{ij}{}^{\mathsf{T}}\boldsymbol{\alpha}$ and consider the following objective function

$$R_{nmq}(\boldsymbol{\delta}, \lambda_m) = \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \Big[\rho_{\tau_k}\Big\{y_{ij} - \mu_{ij\tau_k} - \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_{1\tau_k}/\sqrt{N} - \boldsymbol{z}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_0/\sqrt{m}\Big\} - \rho_{\tau_k}\{y_{ij} - \mu_{ij\tau_k}\}\Big] + \lambda_m\Big(\sum_{k=1}^{q} v_k \sum_{i=1}^{n}|\alpha_i - \delta_{0i}/\sqrt{m}| - |\alpha_i|\Big),$$

The above objective function is a convex function of $\boldsymbol{\delta}$ that is minimized by

$$\widehat{\boldsymbol{\delta}} = \begin{pmatrix} \widehat{\boldsymbol{\delta}}_0 \\ \widehat{\boldsymbol{\delta}}_{1\tau} \end{pmatrix} = \begin{pmatrix} \sqrt{m}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \\ \sqrt{nm}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} - \boldsymbol{\beta}_{\boldsymbol{\tau}}) \end{pmatrix}. \tag{4.21}$$

Our goal is to approximate R_{nmq} by a quadratic function with a unique minimizing value, and use results from Lid Hjort et Pollard (2011) to show that $\widehat{\boldsymbol{\delta}}$ has the same asymptotic distribution of that minimizing value. This quadratic approximation is mainly composed by the Taylor expansion of the expected value and by a linear approximation function.

Let $\widetilde{\boldsymbol{x}_{ij}} = (\boldsymbol{z}_{ij}{}^{\mathsf{T}}, \boldsymbol{x}_{ij}{}^{\mathsf{T}})^{\mathsf{T}}$, $\widetilde{\boldsymbol{\delta}} = (\boldsymbol{\delta}_0{}^{\mathsf{T}}/\sqrt{m}, \boldsymbol{\delta}_{1\tau}{}^{\mathsf{T}}/\sqrt{nm})^{\mathsf{T}}$ and $\varepsilon_{ij\tau} = y_{ij} - \mu_{ij\tau}$. The function $\mathbb{E}(\rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}{}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau}))$ is convex, twice continuously differentiable and reaches its minimum at $\widetilde{\boldsymbol{\delta}} = \mathbf{0}$. It can be represented in the neighbourhood of $\widetilde{\boldsymbol{\delta}} = \mathbf{0}$ as

$$\begin{aligned}\mathbb{E}\left[\rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}{}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau})\right] &= \widetilde{\boldsymbol{\delta}}^{\mathsf{T}}\widetilde{\boldsymbol{x}_{ij}}\,\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau})]\widetilde{\boldsymbol{x}_{ij}}{}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}} \\ &\quad - 2\widetilde{\boldsymbol{\delta}}^{\mathsf{T}}\widetilde{\boldsymbol{x}_{ij}}\,\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau}).\varepsilon_{ij\tau}] + o\left\|\widetilde{\boldsymbol{\delta}}\right\|^2,\end{aligned} \tag{4.22}$$

where $\psi_\tau(\lambda) = \tau - \mathbb{1}(\lambda < 0)$. Since

$$\underset{\boldsymbol{\delta} \in \mathbb{R}^{n+p}}{\operatorname{argmin}} \mathbb{E}\left[\rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau})\right] = \mathbf{0}$$

we have by the first order condition

$$\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau}).\varepsilon_{ij\tau}] = 0, \tag{4.23}$$

and equation (4.22) can be reduced to :

$$\mathbb{E}\left[\rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau})\right] = \widetilde{\boldsymbol{\delta}}^{\mathsf{T}}\widetilde{\boldsymbol{x}_{ij}}\,\mathbb{E}[\psi_\tau(\varepsilon_{ij\tau})]\widetilde{\boldsymbol{x}_{ij}}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}} + o\left\|\widetilde{\boldsymbol{\delta}}\right\|^2. \tag{4.24}$$

The linear approximation function can be considered as a sort of Taylor expansion around $\boldsymbol{\delta} = 0$, see (Pollard, 1991b). Define

$$D_{ij}(\varepsilon_{ij\tau}) = -2\psi_\tau(\varepsilon_{ij\tau}).\varepsilon_{ij\tau}. \tag{4.25}$$

Note that by (4.23), $\mathbb{E}(D_{ij}(\varepsilon_{ij\tau})) = 0$. Define

$$r_{ij}(\widetilde{\boldsymbol{\delta}}) = \rho_\tau(\varepsilon_{ij\tau} - \widetilde{\boldsymbol{x}_{ij}}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau}) - \widetilde{\boldsymbol{\delta}}^{\mathsf{T}}\widetilde{\boldsymbol{x}_{ij}}D_{ij}(\varepsilon_{ij\tau}).$$

Then

$$\begin{aligned} R_{nmq}(\widetilde{\boldsymbol{\delta}}, \lambda_m) = &\sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k\Big(\mathbb{E}\left[\rho_\tau(\varepsilon_{ij\tau_k} - \widetilde{\boldsymbol{x}_{ij}}^{\mathsf{T}}\widetilde{\boldsymbol{\delta}}) - \rho_\tau(\varepsilon_{ij\tau_k})\right]\Big) \\ &+ \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k\widetilde{\boldsymbol{\delta}}^{\mathsf{T}}\widetilde{\boldsymbol{x}_{ij}}D_{ij}(\varepsilon_{ij\tau_k}) + \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k\Big(r_{ij}(\widetilde{\boldsymbol{\delta}}) - \mathbb{E}\left[r_{ij}(\widetilde{\boldsymbol{\delta}})\right]\Big) \\ &+ \lambda_m\Big(\sum_{i=1}^{n}|\alpha_i - \delta_{0i}/\sqrt{m}| - |\alpha_i|\Big). \end{aligned}$$

Using **Lemma** 2.7.0.1, the objective function $R_{nmq}(\widetilde{\boldsymbol{\delta}})$ reduces to

$$
\begin{aligned}
R_{nmq}(\widetilde{\boldsymbol{\delta}}, \lambda_m) &= \widetilde{\boldsymbol{\delta}}^{\mathsf{T}} \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \left(\widetilde{\boldsymbol{x}_{ij}}\, \mathbb{E}[\psi_{\tau_k}(\varepsilon_{ij\tau_k})] \widetilde{\boldsymbol{x}_{ij}}^{\mathsf{T}} \right) \widetilde{\boldsymbol{\delta}} + \widetilde{\boldsymbol{\delta}}^{\mathsf{T}} \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \widetilde{\boldsymbol{x}_{ij}} D_{ij}(\varepsilon_{ij\tau_k}) \\
&\quad + \lambda_m \left(\sum_{i=1}^{n} |\alpha_i - \delta_{0i}/\sqrt{m}| - |\alpha_i| \right) + O\left\| \widetilde{\boldsymbol{\delta}} \right\|^2 + o\left\| \widetilde{\boldsymbol{\delta}} \right\|^2 \\
&= \widetilde{\boldsymbol{\delta}}^{\mathsf{T}} \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \left(\widetilde{\boldsymbol{x}_{ij}}\, \mathbb{E}[\psi_{\tau_k}(\varepsilon_{ij\tau_k})] \widetilde{\boldsymbol{x}_{ij}}^{\mathsf{T}} \right) \widetilde{\boldsymbol{\delta}} + \widetilde{\boldsymbol{\delta}}^{\mathsf{T}} \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \widetilde{\boldsymbol{x}_{ij}} D_{ij}(\varepsilon_{ij\tau_k}) \\
&\quad + \lambda_m \left(\sum_{i=1}^{n} |\alpha_i - \delta_{0i}/\sqrt{m}| - |\alpha_i| \right) + o_p(1) \\
&\simeq \widetilde{\boldsymbol{\delta}}^{\mathsf{T}} \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \left(\widetilde{\boldsymbol{x}_{ij}}\, \mathbb{E}[\psi_{\tau_k}(\varepsilon_{ij\tau_k})] \widetilde{\boldsymbol{x}_{ij}}^{\mathsf{T}} \right) \widetilde{\boldsymbol{\delta}} + \widetilde{\boldsymbol{\delta}}^{\mathsf{T}} \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k \widetilde{\boldsymbol{x}_{ij}} D_{ij}(\varepsilon_{ij\tau_k}) \\
&\quad + \lambda_m \left(\sum_{i=1}^{n} |\alpha_i - \delta_{0i}/\sqrt{m}| - |\alpha_i| \right).
\end{aligned}
\tag{4.26}
$$

By replacing $\widetilde{\boldsymbol{x}_{ij}} = (\boldsymbol{z}_{ij}{}^{\mathsf{T}}, \boldsymbol{x}_{ij}{}^{\mathsf{T}})^{\mathsf{T}}$ and $\widetilde{\boldsymbol{\delta}} = (\boldsymbol{\delta}_0{}^{\mathsf{T}}/\sqrt{m}, \boldsymbol{\delta}_1{}^{\mathsf{T}}/\sqrt{nm})^{\mathsf{T}}$ by their initial value, we have

$$
\begin{aligned}
&R_{nmq}(\boldsymbol{\delta}, \lambda_m) = \\
&- 2\frac{1}{\sqrt{m}} \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k (\boldsymbol{z}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_0 + \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_{1\tau_k}/\sqrt{n}) \psi_{\tau_k}(\varepsilon_{ij\tau_k})(\varepsilon_{ij\tau_k}) \\
&+ \frac{1}{m} \sum_{k=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{m} v_k\, \mathbb{E}[\psi_{\tau_k}(\varepsilon_{ij\tau_k})] (\boldsymbol{z}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_0 + \boldsymbol{x}_{ij}{}^{\mathsf{T}}\boldsymbol{\delta}_{1\tau_k}/\sqrt{n})^2 \\
&+ \lambda_m \left(\sum_{i=1}^{n} |\alpha_i - \delta_{0i}/\sqrt{m}| - |\alpha_i| \right) \\
&= R^{(1)}_{nmq}(\boldsymbol{\delta}) + R^{(2)}_{nmq}(\boldsymbol{\delta}) + R^{(3)}_{mnq}(\boldsymbol{\delta}, \lambda_m).
\end{aligned}
$$

The conditions **A2** and **A3** imply a Lindberg condition for the first term of $R_{nmq}(\boldsymbol{\delta}, \lambda_m)$.

$$R^{(1)}_{nmq}(\boldsymbol{\delta}) = -2\frac{1}{\sqrt{m}}\sum_{k=1}^{q} v_k \left[\boldsymbol{\delta}_0{}^{\mathsf{T}}\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{\tau_k})\boldsymbol{\varepsilon}_{\tau_k} + \frac{\boldsymbol{\delta}_{1\tau_k}{}^{\mathsf{T}}}{\sqrt{n}}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{\tau_k})\boldsymbol{\varepsilon}_{\tau_k}\right]$$
$$\xrightarrow{d} -2\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{B}.$$

$\boldsymbol{B}$ is a zero mean Gaussian vector with covariance matrix $\boldsymbol{D}_0(\boldsymbol{\tau})$.

The second term of $R_{nmq}(\boldsymbol{\delta}, \lambda_m)$ is

$$\begin{aligned}
&R^{(2)}_{nmq}(\boldsymbol{\delta}) \\
&= \frac{1}{m}\sum_{k=1}^{q} v_k \Big[\boldsymbol{\delta}_0{}^{\mathsf{T}}\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{\tau_k})]\boldsymbol{Z}\boldsymbol{\delta}_0 + 2\boldsymbol{\delta}_0{}^{\mathsf{T}}\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{\tau_k})]\boldsymbol{X}\boldsymbol{\delta}_{1\tau_k}/\sqrt{n} \\
&+ \boldsymbol{\delta}_{1\tau_k}{}^{\mathsf{T}}/\sqrt{n}\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_k}(\boldsymbol{\varepsilon}_{\tau_k})]\boldsymbol{X}\boldsymbol{\delta}_1(\tau_k)/\sqrt{n}\Big] \\
&= \frac{1}{m}\boldsymbol{\delta}^{\mathsf{T}}\begin{pmatrix}
\sum_{k=1}^{q} v_k\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_k}]\boldsymbol{Z} & v_1\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_1}]\boldsymbol{X}/\sqrt{n} & \cdots & v_q\boldsymbol{Z}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_q}]\boldsymbol{X}/\sqrt{n} \\
v_1\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_1}]\boldsymbol{Z}/\sqrt{n} & v_1\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_1}]\boldsymbol{X}/n & \cdots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots \\
v_q\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_q}]\boldsymbol{Z}/\sqrt{n} & \mathbf{0} & \cdots & v_q\boldsymbol{X}^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_{\tau_q}]\boldsymbol{X}/n
\end{pmatrix}\boldsymbol{\delta} \\
&\to \boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{D}_1(\boldsymbol{\tau})\boldsymbol{\delta}.
\end{aligned}$$

Finally, using the result of Knight et Fu (2000) the third term of $R_{nmq}(\boldsymbol{\delta}, \lambda_m)$ is

$$R^{(3)}_{nmq}(\boldsymbol{\delta}, \lambda_m) = \frac{\lambda_m}{\sqrt{m}}\sum_{i=1}^{n}\delta_{0i}\operatorname{sign}(\alpha_i) \to \lambda_0\boldsymbol{\delta}_0{}^{\mathsf{T}}\boldsymbol{s}_0,$$

where $\boldsymbol{s}_0{}^{\mathsf{T}} = [\operatorname{sign}(\alpha_i)]_{i=1}^{n}$.

Thus, the limiting form of the objective function is

$$R_{0q}(\boldsymbol{\delta}) = -2\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{B} + \boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{D}_1(\boldsymbol{\tau})\boldsymbol{\delta} + \lambda_0\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{s},$$

where $\boldsymbol{s}^{\mathsf{T}} = (\boldsymbol{s}_0{}^{\mathsf{T}}, \mathbf{0}_{pq}{}^{\mathsf{T}})$. □

Proof of Corollary 4.3.1.1.

Consider again the function R_{mnq} in its matrix form this time

$$\begin{aligned}
R_{mnq}(\boldsymbol{\delta}, \lambda_0) = & -2\frac{\boldsymbol{\delta}_0{}^{\mathsf{T}}}{\sqrt{m}}(\boldsymbol{v} \otimes \boldsymbol{Z})^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau - 2\frac{\boldsymbol{\delta}_{1\tau}{}^{\mathsf{T}}}{\sqrt{nm}}(\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau \\
& + \frac{\boldsymbol{\delta}_0{}^{\mathsf{T}}}{\sqrt{m}}(\boldsymbol{v} \otimes \boldsymbol{Z})^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)](\mathbb{1}_q \otimes \boldsymbol{Z})\frac{\boldsymbol{\delta}_0}{\sqrt{m}} \\
& + 2\frac{\boldsymbol{\delta}_0{}^{\mathsf{T}}}{\sqrt{m}}(\mathbb{1}_q \otimes \boldsymbol{Z})^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)](\boldsymbol{V} \otimes \boldsymbol{X})\frac{\boldsymbol{\delta}_{1\tau}}{\sqrt{nm}} \\
& + \frac{\boldsymbol{\delta}_{1\tau}}{\sqrt{nm}}(\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)](\mathbb{I}_q \otimes \boldsymbol{X})\frac{\boldsymbol{\delta}_{1\tau}}{\sqrt{nm}} \\
& + \lambda_0\boldsymbol{\delta}_0{}^{\mathsf{T}}\boldsymbol{s}_0.
\end{aligned}$$

R_{mnq} is a convex function and twice continuously differentiable. Its first order conditions allows us to derive the expression of $\widehat{\boldsymbol{\delta}}_{1\tau}$ and $\widehat{\boldsymbol{\delta}}_0$. We have

$$\begin{aligned}
(\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)](\mathbb{I}_q \otimes \boldsymbol{X})\frac{\widehat{\boldsymbol{\delta}}_{1\tau}}{\sqrt{nm}} = & (\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau \\
& - (\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)](\mathbb{1}_q \otimes \boldsymbol{Z})\frac{\widehat{\boldsymbol{\delta}}_0}{\sqrt{m}}
\end{aligned}$$

and

$$\begin{aligned}
(\boldsymbol{v} \otimes \boldsymbol{Z})^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)](\mathbb{1}_q \otimes \boldsymbol{Z})\frac{\widehat{\boldsymbol{\delta}}_0}{\sqrt{m}} = & (\boldsymbol{v} \otimes \boldsymbol{Z})^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau - \frac{\lambda_0}{2}\boldsymbol{\delta}_0{}^{\mathsf{T}}\boldsymbol{s}_0 \\
& - (\boldsymbol{v} \otimes \boldsymbol{Z})^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)](\mathbb{I}_q \otimes \boldsymbol{X})\frac{\widehat{\boldsymbol{\delta}}_{1\tau}}{\sqrt{nm}}.
\end{aligned}$$

By replacing the expression of $\widehat{\boldsymbol{\delta}}_0$ in that of $\widehat{\boldsymbol{\delta}}_{1\tau}$ we have

$$\begin{aligned}
\frac{\widehat{\boldsymbol{\delta}}_{1\tau}}{\sqrt{nm}} = & \left[(\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}}\mathbb{E}[\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)]\boldsymbol{M}_{q\boldsymbol{Z}}(\mathbb{I}_q \otimes \boldsymbol{X})\right]^{-1} \\
& \left\{(\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}}\boldsymbol{M}_{q\boldsymbol{Z}}{}^{\mathsf{T}}\boldsymbol{\Psi}_\tau(\boldsymbol{\varepsilon}_\tau)\boldsymbol{\varepsilon}_\tau + \frac{\lambda_0}{2m}(\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}}\boldsymbol{P}_{q\boldsymbol{Z}}{}^{\mathsf{T}}(\boldsymbol{v} \otimes \boldsymbol{Z})^{\mathsf{T}}\boldsymbol{s}_0\right\}.
\end{aligned}$$

In order to complete this proof, we will show the asymptotic normality of the first term of the above expression $(\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}} \boldsymbol{M}_{q\boldsymbol{Z}}{}^{\mathsf{T}} \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{\boldsymbol{\tau}}) \boldsymbol{\varepsilon}_{\boldsymbol{\tau}}$ which is a vector of length pq. As explained by Barry *et al.* (2018b), we can show the result for a single τ. Let $\boldsymbol{X}_{\boldsymbol{Z}} = \boldsymbol{M}_{q\boldsymbol{Z}} \boldsymbol{X}$ then we have

$$m^{-1} \boldsymbol{X}_{\boldsymbol{Z}}{}^{\mathsf{T}} \boldsymbol{\Psi}_{\tau}(\boldsymbol{\varepsilon}_{\tau}) \boldsymbol{\varepsilon}_{\tau} = \sum_{i=1}^{n} \sum_{j=1}^{m} \boldsymbol{x}_{ij\boldsymbol{Z}} \psi_{\tau}(\varepsilon_{ij\tau}) \varepsilon_{ij\tau} / m.$$

To demonstrate the asymptotic normality we apply the Cramér-Wold device and verify the Lyapunov's condition. Let $T_{Ni} = \boldsymbol{\eta}^{\mathsf{T}} (\boldsymbol{V} \otimes \boldsymbol{X}_i)^{\mathsf{T}} \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}) \boldsymbol{\varepsilon}_{i\boldsymbol{\tau}}$ and consider $N^{-1/2} \sum_{i=1}^{n} T_{Ni}$, where $\boldsymbol{\eta}$ is a $pq \times 1$ unit vector, $\boldsymbol{\eta}^{\mathsf{T}} \boldsymbol{\eta} = 1$. The summands T_{Ni} are independent with $\mathbb{E}[T_{Ni}] = 0$ and $\text{Var}\left[n^{-1/2} \sum_{i=1}^{n} T_{Ni}\right] > \nu\prime > 0$, by condition **A2'**. When we apply Minkowski's inequality, we have

$$\begin{aligned}
\mathbb{E}|T_{Ni}|^{2+\nu} &= \mathbb{E} \left| \sum_{l=1}^{q} \sum_{j=1}^{m} \sum_{k=1}^{p} m^{-1} \eta_{kl} v_l x_{ij\boldsymbol{Z}}^{k} \psi_{\tau_l}(\varepsilon_{ij\tau_l}) \varepsilon_{ij\tau_l} \right|^{2+\nu} \\
&\leq \left[\sum_{l=1}^{q} \sum_{j=1}^{m} \sum_{k=1}^{p} \left(\mathbb{E} \left| m^{-1} \eta_{kl} v_l x_{ij\boldsymbol{Z}}^{k} \psi_{\tau_l}(\varepsilon_{ij\tau_l}) \varepsilon_{ij\tau_l} \right|^{2+\nu} \right)^{\frac{1}{2+\nu}} \right]^{2+\nu} \\
&\leq \left[\sum_{l=1}^{q} \sum_{j=1}^{m} \sum_{k=1}^{p} m^{-1} |\eta_{kl} v_l x_{ij\boldsymbol{Z}}^{k}| \left(\mathbb{E} \left| \psi_{\tau_l}(\varepsilon_{ij\tau_l}) \varepsilon_{ij\tau_l} \right|^{2+\nu} \right)^{\frac{1}{2+\nu}} \right]^{2+\nu \cdot} \\
&\leq M \Delta p^{1+\nu},
\end{aligned}$$

where the last inequality follows by $\mathbb{E}|\psi_{\tau}(\varepsilon_{ij\tau})|^{4+\nu} < \Delta$ and $\mathbb{E}|\varepsilon_{ij\tau}|^{4+\nu} < \Delta$.

Then by the Liapounov CLT $N^{-1/2} (\boldsymbol{V} \otimes \boldsymbol{X})^{\mathsf{T}} \boldsymbol{M}_{q\boldsymbol{Z}}{}^{\mathsf{T}} \boldsymbol{\Psi}_{\boldsymbol{\tau}}(\boldsymbol{\varepsilon}_{\boldsymbol{\tau}}) \boldsymbol{\varepsilon}_{\boldsymbol{\tau}}$ is Gaussian and by condition **A2'**, $\widehat{\boldsymbol{\delta}}_{1\boldsymbol{\tau}}$ is zero mean Gaussian vector with covariance matrix $\text{Var}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}}) = \widetilde{\boldsymbol{D}}_{1q}^{-1}(\boldsymbol{\tau}) \widetilde{\boldsymbol{D}}_{0q}(\boldsymbol{\tau}) \widetilde{\boldsymbol{D}}_{1q}^{-1}(\boldsymbol{\tau})$. □

Proof of Corollary 4.3.1.2.

The proof of **Corollary**4.3.1.2 follows immediately from the proof of **Corollary**4.3.1.1, with a single τ. □

CONCLUSION

Dans la présente thèse, nous avons introduit une nouvelle classe d'estimateurs pour l'analyse des données longitudinales. Cette nouvelle classe, basée sur la régression des moindres carrés asymétriques pondérés (régression expectile), est une généralisation de l'estimateur classique des moindres carrés ordinaires. En plus d'estimer un effet moyen des régresseurs, elle estime leur influence sur tout autre percentile de la distribution conditionnelle de la variable réponse. Parmi les modèles d'analyse des données longitudinales, nous avons étudié la généralisation du modèle GEE et du modèle EF qui sont deux modèles couramment utilisés en biostatistique et en économétrie. Les trois articles de la présente thèse réunissent la généralisation de ces modèles.

Le premier article de la présente thèse est soumis au journal «Computational Statistics and Data Analysis» et aborde la généralisation du modèle GEE. Le modèle GEE est un modèle marginal très populaire pour l'analyse des données longitudinales. Le modèle GEE modélise la moyenne marginale et tient compte de la dépendance intra-groupe des données provenant d'un même sujet par l'inclusion dans le modèle d'une structure de corrélation hypothétique dite «working correlation structure». L'adaptation de la régression expectile au modèle GEE a donné naissance à une nouvelle classe d'estimateurs. Cette nouvelle classe, que nous avons désignée GEEE, permet au modèle GEE d'étudier l'influence des régresseurs sur la localisation, l'échelle et la forme de la distribution de la variable réponse. Dans notre article, nous avons dérivé

les propriétés asymptotiques de l'estimateur GEEE et avons suggéré un estimateur robuste et convergent pour sa matrice de variance-covariance. Nous avons présenté un algorithme montrant la généralisation des structures de corrélation existantes pour le modèle GEE.

Nous avons mené des simulations exhaustives pour évaluer la performance des estimateurs GEEE. Les données ont été générées selon deux scénarios (location-shift et location-scale-shift) avec trois distributions $\left(\mathcal{N}(0,1), t_3, \chi^3_2\right)$. La dépendance intra-groupe des données provient d'une structure de corrélation AR1 obtenue à l'aide du package Copula (Hofert *et al.*, 2017) de R (R Core Team, 2018). Nous avons ajusté le modèle GEEE aux données simulées avec quatre structures de corrélation différentes : indépendante, échangeable, AR1 et non-structurée, ainsi que le modèle mixte de la régression quantile linéaire «LQMM» (Geraci et Bottai, 2007, 2014). Les résultats de la simulation ont montré les qualités favorables et les avantages de l'estimateur GEEE dans les deux scénarios. Nous avons utilisé le critère QIC de Pan (2001) pour discriminer parmi les quatre structures de corrélation. Le QIC est plus susceptible de sélectionner la corrélation (AR1) parmi les quatre structures de corrélation sélectionnées dans la simulation.

Finalement, nous avons ajusté le modèle GEEE aux données sur les douleurs liées au travail à l'accouchement pour étudier l'effet d'un nouveau traitement. Les résultats ont révélé une forte corrélation entre le traitement et le temps. Les résultats ont montré également que l'évolution de la douleur en fonction du temps varie selon qu'on se trouve au centre ou à gauche ou à droite de la queue de la distribution de la variable réponse.

Il faut noter que la régression quantile (RQ) a été adaptée au modèle GEE. Chen *et al.* (2004) ont proposé un estimateur RQ pour les données corrélées en utilisant l'approche GEE basée sur la structure de corrélation indépendante. Fu et Wang (2012) ont combiné la RQ et les fonctions d'estimation intra et inter pour tenir compte de la dépendance des données longitudinales. Lu et Fan (2015) ont adapté la RQ au modèle GEE avec une structure d'autocorrélation stationnaire. Malgré la robustesse du modèle RQ, sa combinaison avec l'approche GEE est difficile. L'adaptation des structures de corrélation dans le cadre du modèle RQ n'est pas garantie pour tous les types de structures, (Farcomeni et Marino, 2015).

Dans le second article, qui a été soumis au journal «Econometric Reviews», nous avons considéré l'adaptation de la RE au modèle linéaire avec effets-fixes (EF). Le modèle EF est un modèle couramment utilisé pour l'analyse des données longitudinales, particulièrement en économétrie. Notamment, parce qu'elle tient compte, implicitement, de la corrélation entre les régresseurs du modèle et les caractéristiques individuelles non-observées. Sa combinaison avec la RE a donné lieu au modèle de la régression expectile avec effets-fixes, que nous avons désigné par ERFE. Le modèle ERFE, en plus de contrôler les effets-individuels, capture aussi l'hétérogénéité des effets des régresseurs et tient compte de l'hétérogénéité non-observée.

Comme dans le cadre du modèle EF, le modèle ERFE est confronté au problème de la dimension du paramètre individuel, qui croît au fur et à mesure que la taille de l'échantillon augmente. Ce problème, communément appelé «incidental parameter problem» complique l'estimation des paramètres du modèle ERFE. Pour contourner le problème, nous avons pu appliquer la méthode de la «within-transformation» de manière itérative pour estimer les paramètres du modèle ERFE. Dans notre algo-

rithme, nous avons éliminé le paramètre individuel par l'application d'une matrice de projection aux données initiales du modèle. Ensuite, nous avons appliqué l'algorithme des moindres carrés pondérés itératifs aux données transformées, comme dans le cadre de la régression expectile simple, jusqu'à l'obtention de la convergence. Nous avons montré les propriétés asymptotiques de l'estimateur ERFE et nous avons proposé un estimateur robuste et convergent de sa matrice de variance-covariance pour l'inférence.

Nous avons généré les données de simulation selon les deux scénarios (location-shift et location-scale-shift) avec les trois distributions $\left(\mathcal{N}(0,1), t_3, \chi^3_2\right)$ pour l'évaluation de la performance de l'estimateur ERFE. Nous avons comparé notre estimateur à celui du modèle de la régression quantile avec effets-fixes (QRFE), (Koenker, 2004). Les résultats des simulations sont mitigés. Dans un premier temps, les résultats montrent que l'estimateur ERFE est plus performant que l'estimateur QRFE lorsque les données simulées proviennent du scénario «location-shift». Tandis que dans un scénario «location-scale-shift», l'estimateur ERFE performe moins bien que l'estimateur QRFE.

Nous avons illustré l'application du modèle ERFE sur des données réelles pour étudier le rendement scolaire sur le salaire. Les données proviennent du «Panel Study of Income Dynamic (PSID)» de Baltagi et Khanti-Akom (1990). Les données PSID sont fréquemment employées dans les manuels économétriques (Greene, 2012; Baltagi, 2008; Wooldridge, 2002; Cameron et Trivedi, 2008) pour comparer les méthodes d'analyse des données longitudinales. Les résultats de notre article montrent l'avantage de l'utilisation du modèle ERFE sur le modèle EF, par exemple. Cependant, nous avons remarqué une différence d'échelle entre les résultats du modèle ERFE et

ceux du modèle QRFE pour certains régresseurs. Cette différence n'est pas surprenante lorsque nous considérons la distribution de ces régresseurs et les caractéristiques de l'estimateur ERFE. D'un côté, ces régresseurs qui incluent le statut marital et l'éthnicité ont une variation intra-groupe quasiment nulle et donc, sont presque des régresseurs invariants dans le temps. D'un autre côté, l'estimateur ERFE est un estimateur intra-groupe dont l'expression est fonction de la différence des observations et leur moyenne intra-groupe. Par conséquent, il est attendu que l'estimateur ERFE de ces régresseurs soit proche de zéro. Nous avons également sorti les résultats des modèles de la RE et de la RQ qui montrent que les deux modèles sont sur la même échelle et révèlent la même tendance.

Le modèle ERFE améliore le modèle EF qui correspond au modèle ERFE de niveau $\tau = 0.5$. Cependant, le modèle ERFE a le même défaut que le modèle EF. Dans le cadre du modèle ERFE, la transformation des données élimine le paramètre individuel, mais également tous les régresseurs invariants dans le temps. Ainsi, il n'est pas possible de réaliser de l'inférence sur les coefficients des régresseurs invariants dans le temps avec le modèle ERFE.

Il faut noter que la méthode de transformation «within-transformation», qui permet d'éliminer le paramètre individuel du modèle EF, ne s'applique pas dans le cadre de la RQ. À la place, Koenker (2004) a fait appel à la représentation de Bahadur pour atténuer la question du «incidental parameter problem».

Le troisième et dernier article porte aussi sur le modèle EF. Toutefois, il adopte une approche différente pour palier au «incidental parameter problem». Nous introduisons, dans cet article, le modèle ERFE avec pénalité (PERFE) où la pénalité

est appliquée au paramètre des effets individuels. Ce modèle s'inspire de l'approche d'Henderson (1950) qui a présenté l'estimateur BLUE comme un estimateur avec pénalité. Au lieu d'appliquer la pénalité l_2, nous avons choisi la pénalité l_1 dont les propriétés conviennent mieux à nos besoins. La pénalité l_1 permet de réaliser simultanément l'estimation des paramètres et la régularisation (rétrécissement) du paramètre individuel. En plus de proposer des estimateurs plus efficaces et une solution «clairsemée» (sparse) des effets individuels, le modèle PERFE permet l'estimation des coefficients des régresseurs invariants dans le temps. Cette dernière caractéristique distingue le modèle PERFE comme modèle alternatif et pertinent pour le modèle ERFE. Il faut noter que le modèle ERFE est équivalent au modèle PERFE lorsque le paramètre de régularisation tend vers zéro. Ainsi, nous pouvons utiliser le modèle PERFE avec un paramètre de régularisation proche de zéro pour estimer un modèle ERFE qui permet l'inférence sur les régresseurs invariants dans le temps.

La composante de la fonction objectif qui introduit la pénalité n'est pas différentiable en zéro par rapport au paramètre individuel. Pour cette raison, les algorithmes classiques, comme celui de Gauss-Newton, ne s'appliquent plus. Nous avons proposé un «block relaxation algorithm» combiné à un «coordinate descent algorithm» pour efficacement estimer le modèle PERFE et régulariser le paramètre individuel. Nous avons utilisé une astuce, empruntée à Mkhadri *et al.* (2017), pour réduire la grille du paramètre de régularisation. Cette astuce s'appuie sur les conditions de Karush–Kuhn–Tucker (KKT) pour obtenir la valeur maximale du paramètre de régularisation $\lambda_{\max}$ qui rétrécit toutes les valeurs du vecteur individuel à 0. L'estimation du modèle PERFE est réalisée en estimant le modèle pour K valeurs du paramètre de régularisation choisi entre $\lambda_{\max}$ et $10^{-4}\lambda_{\max}$. Le choix de la valeur de K est arbitraire. Nous avons choisi $K = 10$ pour la simulation et $K = 100$ pour l'estimation

des données réelles. Le modèle correspondant au paramètre de régularisation optimal a été retenu. Le paramètre optimal et le modèle optimal ont été sélectionnés en fonction des valeurs du critère d'information bayésien (BIC). Nous avons également dérivé les propriétés asymptotiques de l'estimateur PERFE.

Les simulations exhaustives montrent que l'estimateur PERFE réduit le biais et améliore l'efficacité. Les résultats de la simulation ont montré sa performance dans divers scénarios et ses avantages par rapport à l'estimateur ERFE et l'estimateur «penalized quantile regression with fixed-effects (PQRFE)» de Koenker (2004). En effet, peu importe le scénario «location-shift» ou «location-scale-shift» et la distribution $\left(\mathcal{N}(0,1), t_3, \text{ et } \chi^3_2\right)$, l'estimateur PERFE a toujours le plus faible biais et la meilleure efficacité que les estimateurs ERFE et PQRFE.

Nous avons repris les données réelles (PSID) du second article pour illustrer le modèle PERFE sur des données réelles. Avant tout, il faut noter l'estimation des effets des régresseurs invariants dans le temps par le modèle PERFE permettant ainsi l'inférence sur ces régresseurs.

Les résultats des deux modèles PERFE et PQRFE révèlent l'existence de l'hétérogénéité des effets des régresseurs et en particulier l'effet de l'éducation sur le salaire. Les résultats des deux modèles ont généralement la même échelle, mais on observe parfois des différences. Par exemple, c'est le cas pour l'effet de l'éducation sur le salaire, où le résultat du modèle PQRFE est très similaire à celui du modèle RQ et du modèle RE classique.

Eilers (2013) fait valoir que les statistiques asymétriques (quantile et expectile) ont

beaucoup à apporter à la recherche statistique théorique et empirique et à la modélisation des données, en particulier. Avec elles, les chercheurs sont mieux équipés pour capturer l'hétérogénéité des effets des regresseurs et pour étudier l'hétérogénéité non-observée.

Aujourd'hui, grâce à la grande puissance computationnelle des ordinateurs, la RQ est devenue un outil de modélisation très réputé en statistique et en économétrie. Néanmoins, son déploiement n'est pas intégral. Il existe d'autres branches de la modélisation où la RQ s'implante trop peu. C'est le cas de la modélisation des variables réponses discrètes, où l'application de la RQ est délicate et non systématique. Nous pensons à l'application de la RQ au modèle de comptage (Machado et Silva, 2005), ou au modèle GEE avec la difficulté de généraliser et d'adapter des structures de corrélations sophistiquées dans le contexte de la RQ (Farcomeni et Marino, 2015). Parmi les obstacles potentiels d'implémentation systématique mentionnons que sa fonction de perte n'est pas différentiable en zéro. À cause de cette limite, lors de l'analyse des données avec la méthode RQ, le logiciel peut émettre des avertissements dénonçant que les résultats obtenus ne sont pas uniques, sans pour autant savoir comment procéder. Un autre désavantage réside dans le fait que malgré l'excellente puissance computationnelle, l'application de la RQ est coûteuse et demeure un défi computationnel de taille, surtout avec les données de hautes dimensions.

Par ailleurs, la RE est basée sur une fonction de perte convexe et continûment différentiable et sa minimisation aboutit à une solution unique. Son implémentation avec l'algorithme itératif de la pondération des moindres carrés ne pose pas de problèmes numériques et s'exécute avec un temps de calcul relativement faible. Contrairement à certains quantiles, l'expectile est bien défini pour des distributions discrètes (Eilers,

2013) et cette ouverture peut être exploitée pour une combinaison, par exemple, du modèle RE et du modèle GEE avec une variable réponse dichotomique ou de comptage. Ce développement pourrait mener à de nouveaux enseignements sur l'étude de la sur-dispersion dans le modèle de Poisson, par exemple.

Dans la présente thèse, nous avons contribué à la généralisation de la RE en l'adoptant à certains modèles statistiques pour l'analyse des données longitudinales. Les deux modèles, la RQ et la RE, permettent aux chercheurs de capturer l'hétérogénéité des effets et de mieux visualiser la dynamique de la relation entre les régresseurs et la variable réponse. Cependant, comme à l'heure actuelle l'expansion de la RQ se bute à des obstacles techniques et computationnels, nous suggérons de mettre l'accent sur le développement de la RE. Nous suggérons de poursuivre l'adaptation de la RE aux modèles statistiques précédemment développés dans le cadre de la régression des moindres carrés ordinaires. Il est souhaitable de travailler à l'élaboration de packages et de l'implémenter dans divers logiciels standards de statistiques. Nous concluons que lorsque l'objectif est d'estimer des effets sur la moyenne conditionnelle, la RE devrait toujours être privilégiée.

Appendices

APPENDICE I

SOME ASYMPTOTIC RESULTS

In this section, we list some important results that we used to demonstrate most of the **Theorems** in our thesis.

We first present **Corollary A.1** and **Theorem A.1** of Hjort and Pollard Lid Hjort et Pollard (2011). **Corollary A.1** and **Theorem A.1** are results on the asymptotic properties of the estimators defined by minimization of convex criterion functions. **Corollary A.1** is a general results whereas **Theorem A.1** gives the criteria of convergence for an independent sample.

Corollary A.1 (Basic Corollary of Lid Hjort et Pollard (2011)). *Consider* $A_n(\boldsymbol{s})$ *is convex and can be represented as* $\frac{1}{2}\boldsymbol{s}^T\boldsymbol{V}\boldsymbol{s} + \boldsymbol{u}_n{}^T\boldsymbol{s} + c_n + r_n(\boldsymbol{s})$, *where* $\boldsymbol{V}$ *is symmetric and positive definite,* $\boldsymbol{u}_n$ *is stochastically bounded,* c_n *is arbitrary, and* $r_n(\boldsymbol{s})$ *goes to zero in probability for each* $\boldsymbol{s}$. *Then* $\boldsymbol{\alpha}_n$ *the argmin of* A_n, *is only* $o_p(1)$ *away from* $\boldsymbol{\beta}_n = -\boldsymbol{V}^{-1}\boldsymbol{u}_n$, *the argmin of* $\frac{1}{2}\boldsymbol{s}^T\boldsymbol{V}\boldsymbol{s} + \boldsymbol{u}_n{}^T\boldsymbol{s} + c_n$. *If also* $\boldsymbol{u}_n \xrightarrow{d} \boldsymbol{u}$ *then* $\boldsymbol{\alpha}_n \xrightarrow{d} -\boldsymbol{V}^{-1}\boldsymbol{u}$.

Let $Y_1, \ldots, Y_n$ be independent random variables with different density functions $f_i(y_i, \boldsymbol{\theta}_0, \eta_i)$, where $\boldsymbol{\theta}_0$ is a p-dimensional parameter of interest. Let $g_i(y_i, \boldsymbol{\theta})$ be convex

functions in $\boldsymbol{\theta}$, and $\widehat{\boldsymbol{\theta}}$ an estimator of $\boldsymbol{\theta}_0$ which minimizes $\sum_{i\leq n} g_i(y_i, \boldsymbol{\theta})$. Suppose that

$$g_i(y_i, \boldsymbol{\theta}_0 + \boldsymbol{t}) - g_i(y_i, \boldsymbol{\theta}_0) = D_i(y_i)^\intercal \boldsymbol{t} + R_i(y_i, \boldsymbol{t}),$$

where $\mathbb{E}[D_i(Y_i)] = 0$. Write $\boldsymbol{B}_i = \mathrm{Var}[D_i(Y_i)]$ and

$$\mathbb{E}[R_i(Y_i, \boldsymbol{t})] = \frac{1}{2}\boldsymbol{t}^\intercal \boldsymbol{A}_i \boldsymbol{t} + v_{i,0}(\boldsymbol{t}) \text{ and } \mathrm{Var}[R_i(Y_i, \boldsymbol{t})] = v_i(\boldsymbol{t}).$$

Then,

Theorem A.1 (Theorem 2.2 of Lid Hjort et Pollard (2011)). *Assume that* $\sum_{i\leq n} v_{i,0}(\boldsymbol{s}/\sqrt{n}) \to 0$ *and* $\sum_{i\leq n} v_i(\boldsymbol{s}/\sqrt{n}) \to 0$ *for each* $\boldsymbol{s}$, *and that* $\boldsymbol{J}_n/n = \sum_{i\leq n} \boldsymbol{A}_i/n$ *and* $\boldsymbol{K}_n/n = \sum_{i\leq n} \boldsymbol{B}_i/n$ *converge to* $\boldsymbol{J}$ *and* $\boldsymbol{K}_n$, *where* $\boldsymbol{J}$ *is positive definite. Then* $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ *is only* $o_p(1)$ *away from* $-\boldsymbol{J}^{-1} n^{-1/2} \sum_{i\leq n} D_i(Y_i)$. *If in particular the Lindeberg requirements are fulfilled for the* $D_i(Y_i)$ *sequence, then* $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{J}^{-1}\boldsymbol{K}\boldsymbol{J}^{-1}\right)$.

In the following, we present the Cramér-Wold device and the Lyapunov's CLT. Two useful results for showing the asymptotic normality of a vector of random variables.

Theorem A.2 (Theorem 5.1 of White (2001)). ***(Cramér-Wold device).*** *Let* $\{\mathbf{b}_n\}$ *be a sequence of* $k \times 1$ *random vectors and suppose that for any* $k \times 1$ *real vector* $\boldsymbol{\lambda}$ *such that* $\boldsymbol{\lambda}^\intercal \boldsymbol{\lambda} = 1$, $\boldsymbol{\lambda}^\intercal \mathbf{b}_n \xrightarrow{d} \boldsymbol{\lambda}^\intercal \boldsymbol{z}$, *where* $\boldsymbol{z}$ *is a* $k \times 1$ *vector with joint distribution function* F. *Then the limiting distribution function of* $\mathbf{b}_n$ *exists and equals* F.

Theorem A.3 (Theorem 5.11 of White (2001)). *Let* $\{Z_{nt}\}$ *be a sequence of independent random scalars with* $\mu_{nt} = \mathbb{E}[Z_{nt}]$, $\sigma_{nt}^2 = \mathrm{Var}[Z_{nt}]$, *and* $\mathbb{E}|Z_{nt}|^{2+\nu} < \Delta < \infty$ *for some* $\nu > 0$ *and all* n *and* t. *Define* $\overline{Z}_n = n^{-1}\sum_{t=1}^n Z_{nt}$, $\overline{\mu}_n = n^{-1}\sum_{t=1}^n \mu_{nt}$ *and* $\overline{\sigma}_n^2 = \mathrm{Var}[\sqrt{n}\overline{Z}_n] = n^{-1}\sum_{t=1}^n \sigma_{nt}^2$. *If* $\overline{\sigma}_n^2 > \nu\prime > 0$ *for all* n *sufficiently large, then* $\sqrt{n}(\overline{Z}_n - \overline{\mu}_n)/\overline{\sigma}_n \xrightarrow{d} \mathcal{N}(0,1)$.

The following result is a Corollary of the Markov's law of large numbers (LLN) that will be invoked to show the convergence of the estimators of the covariance matrices in our proposed expectile regression models.

Corollary A.1 (Corollary 3.9 of White (2001)). *Let* $\{Z_t\}$ *be a sequence of independent random variables such that* $\mathbb{E}|Z_t|^{1+\nu} < \Delta < \infty$ *for some* $\nu > 0$ *and all* t. *Then* $\overline{Z}_n - \overline{\mu}_n \xrightarrow{p} 0$, *where* $\overline{Z}_n = n^{-1}\sum_{t=1}^{n} Z_t$ *and* $\overline{\mu}_n = n^{-1}\sum_{t=1}^{n}\mathbb{E}[Z_t]$.

We finish by an other result useful in the proof of the convergence of the variances-covariance matrix estimators.

Lemma A.1 (Lemma A.1 of Hansen (2007)).
For matrices $\boldsymbol{A}$ *and* $\boldsymbol{B}$, $\mathbb{E}\,\|\boldsymbol{A}\otimes\boldsymbol{B}\|^r \leq (E\,\|\boldsymbol{A}\|^{2r}\,E\,\|\boldsymbol{B}\|^{2r})^{1/2}$.

RÉFÉRENCES

Abdous, B. et Remillard, B. (1995). Relating quantiles and expectiles under weighted-symmetry. *Annals of the Institute of Statistical Mathematics*, *47*(2), 371–384. http://dx.doi.org/10.1007/BF00773468. Récupéré de https://doi.org/10.1007/BF00773468

Aigner, D., Amemiya, T. et Poirier, D. (1976). On the estimation of production frontiers : Maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review*, *17*(2), 377.

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*, *49*(4), 431–34.

Arellano, M. (2003). *Panel Data Econometrics.* Oxford University Press.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*(2), 171–178.

Baltagi, B. (2008). *Econometric Analysis of Panel Data.* John Wiley & Sons.

Baltagi, B. et Khanti-Akom, S. (1990). On efficient estimation with panel data : An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics*, *5*(4), 401–06.

Barrodale, R. F. D. K. (1973). An improved algorithm for discrete l_1 linear approximation. *SIAM Journal on Numerical Analysis*, *10*(5), 839–848.

Barry, A., Oualkacha, K. et Charpentier, A. (2018a). *Weighted asymmetric least squares regression for longitudinal data using GEE.* Department of Mathematics, Université du Québec à Montréal. Unpublished manuscript.

Barry, A., Oualkacha, K. et Charpentier, A. (2018b). *Weighted asymmetric least squares regression for longitudinal data with fixed-effects.* Department of Mathematics, Université du Québec à Montréal. Unpublished manuscript.

Bickel, P. J. (1975). One-step huber estimates in the linear model. *Journal of the American Statistical Association, 70*(350), 428–434.

Cameron, A. et Trivedi, P. (2005). *Microeconometrics.* Cambridge University Press.

Cameron, A. C. et Trivedi, P. K. (2008). *Microeconometrics Using Stata* (1st éd.). Stata Press.

Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal, 14*(3), 368–386. `http://dx.doi.org/10.1111/j.1368-423X.2011.00349.x`

Card, D. (1996). The effect of unions on the structure of wages : A longitudinal analysis. *Econometrica, 64*(4), 957–79.

Chen, L., Wei, L.-J. et Parzen, M. I. (2004). *Quantile Regression for Correlated Observations*, Dans D. Y. Lin et P. J. Heagerty (dir.). *Proceedings of the Second Seattle Symposium in Biostatistics : Analysis of Correlated Data*, (p. 51–69). Springer New York : New York, NY

Cornwell, C. et Rupert, P. (1988). Efficient estimation with panel data : An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics, 3*(2), 149–55.

Davis, C. S. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in Medicine*, *10*(12), 1959–1980. http://dx.doi.org/10.1002/sim.4780101210

de Leeuw, J. (1994). Block-relaxation algorithms in statistics. Dans H.-H. Bock, W. Lenski, et M. M. Richter (dir.). *Information Systems and Data Analysis*, 308–324., Berlin, Heidelberg. Springer Berlin Heidelberg.

Dieleman, J. L. et Templin, T. (2014). Random-effects, fixed-effects and the within-between specification for clustered data in observational health studies : A simulation study. *PLOS ONE*, *9*(10), 1–17. http://dx.doi.org/10.1371/journal.pone.0110257

Diggle, P., Heagerty, P., Liang, K. et Zeger, S. (2013). *Analysis of Longitudinal Data*. Oxford Statistical Science Series. OUP Oxford.

Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, *1*(1), 93–125.

Eilers, P. H. (2013). Discussion : The beauty of expectiles. *Statistical Modelling*, *13*(4), 317–322.

Farcomeni, A. et Marino, M. F. (2015). Linear quantile regression models for longitudinal experiments : an overview. *METRON*, *73*(2), 229–247.

Farooq, M. et Steinwart, I. (2017). An svm-like approach for expectile regression. *Computational Statistics and Data Analysis*, *109*, 159–181. http://dx.doi.org/10.1016/j.csda.2016.11.010

Fitzenberger, B. et Wilke, R. A. (2006). Using quantile regression for duration analysis. *Allgemeines Statistisches Archiv*, *90*(1), 105–120. http://dx.doi.org/10.1007/s10182-006-0224-2

Fitzmaurice, G., Davidian, M., Verbeke, G. et Molenberghs, G. (2008). *Longitudinal Data Analysis*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press.

Frees, E. (2004). *Longitudinal and Panel Data : Analysis and Applications in the Social Sciences*. Cambridge University Press.

Friedman, J., Hastie, T. et Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, Articles*, *33*(1), 1–22. http://dx.doi.org/10.18637/jss.v033.i01

Fu, L. et Wang, Y.-G. (2012). Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, *56*(8), 2526 – 2538. http://dx.doi.org/https://doi.org/10.1016/j.csda.2012.02.005

Furlotte, N. A., Eskin, E. et Eyheramendy, S. (2012). Genome-wide association mapping with longitudinal data. *Genetic Epidemiology*, *36*(5), 463–471.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, 246–263.

Galvao, A. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, *164*(1), 142–157. cited By (since 1996)9.

Galvao, A. et Montes-Rojas, G. (2010). Penalized quantile regression for dynamic panel data. *Journal of Statistical Planning and Inference*, *140*(11), 3476–3497. cited By (since 1996)5.

Gardiner, J. C., Luo, Z. et Roman, L. A. (2009). Fixed effects, random effects and gee : What are the differences ? *Statistics in Medicine*, *28*(2), 221–239. http://dx.doi.org/10.1002/sim.3478. Récupéré de https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3478

Geraci, M. et Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, *8*(1), 140–154.

Geraci, M. et Bottai, M. (2014). Linear quantile mixed models. *Statistics and Computing*, *24*(3), 461–479. http://dx.doi.org/10.1007/s11222-013-9381-9

Greene, W. H. (2011). *Econometric analysis* (7th ed.. éd.). Upper Saddle River, N.J. : Prentice Hall.

Greene, W. H. (2012). *Econometric analysis* (7th ed.. éd.). Boston : Montréal : Prentice Hall.

Gu, Y. et Zou, H. (2016). High-dimensional generalizations of asymmetric least squares regression and their applications. *Ann. Statist.*, *44*(6), 2661–2694. http://dx.doi.org/10.1214/15-AOS1431. Récupéré de https://doi.org/10.1214/15-AOS1431

Hansen, C. B. (2007). Asymptotic properties of a robust variance matrix estimator for panel data when t is large. *Journal of Econometrics*, *141*(2), 597 – 620.

Hardin, J. et Hilbe, J. (2012). *Generalized Estimating Equations, Second Edition.* Chapman & Hall/CRC.

Harding, M. et Lamarche, C. (2009). A quantile regression approach for estimating panel data models using instrumental variables. *Economics Letters*, *104*(3), 133–135. cited By (since 1996)7.

Hastie, T., Tibshirani, R. et Friedman, J. (2001). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction.* Springer series in statistics. Springer.

Henderson, C. (1950). Estimation of genetic parameters (abstract). *Ann. Math. Statist.*, *21*, 309–310.

Hofert, M., Kojadinovic, I., Maechler, M. et Yan, J. (2017). *copula : Multivariate Dependence with Copulas.* R package version 0.999-18.

Hsiao, C. (2007). Panel data analysis—advantages and challenges. *TEST*, *16*(1), 1–22.

Hsiao, C. (2015). *Analysis of Panel Data.* Cambridge University Press.

Hui, Z. et Trevor, H. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, *67*(2), 301–320. `http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x`

Hunter, D. R. et Lange, K. (2000). Quantile regression via an mm algorithm. *Journal of Computational and Graphical Statistics*, *9*(1), 60–77.

Jang, M. (2011). Working correlation selection in generalized estimating equations.

Jiang, C., Jiang, M., Xu, Q. et Huang, X. (2017). Expectile regression neural network model with applications. *Neurocomputing*, *247*, 73–86. `http://dx.doi.org/10.1016/j.neucom.2017.03.040`

Jung, S.-H. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association*, *91*(433), 251–257.

Kim, M. et Lee, S. (2016). Nonlinear expectile regression with application to value-at-risk and expected shortfall estimation. *Computational Statistics and Data Analysis*, *94*, 1–19. http://dx.doi.org/10.1016/j.csda.2015.07.011

Kneib, T. (2013a). Beyond mean regression. *Statistical Modelling*, *13*(4), 275–303. http://dx.doi.org/10.1177/1471082X13494159

Kneib, T. (2013b). Rejoinder. *Statistical Modelling*, *13*(4), 373–385. http://dx.doi.org/10.1177/1471082X13494531

Kneib, T. (2013c). Rejoinder. *Statistical Modelling*, *13*(4), 373–385.

Knight, K. et Fu, W. (2000). Asymptotics for Lasso-Type Estimators. *The Annals of Statistics*, *28*(5), 1356–1378.

Kocherginsky, M., He, X. et Mu, Y. (2005). Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics*, *14*(1), 41–55.

Koenker, Roger, .-e., Chernozhukov, Victor, e., He, Xuming, e. et Peng, Limin, e. (2018). *Handbook of quantile regression.* Boca Raton : CRC Press, Taylor Francis Group. "A Chapman Hall book.".

Koenker, R. (1993). When are expectiles percentiles? *Econometric Theory*, *9*(3), 526–527.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, *91*(1), 74–89. http://dx.doi.org/10.1016/j.jmva.2004.05.006

Koenker, R. (2005). *Quantile regression*, volume 38 de *Econometric Society Monographs.* Cambridge University Press.

Koenker, R. (2013). Discussion : Living beyond our means. *Statistical Modelling*, *13*(4), 323–333. http://dx.doi.org/10.1177/1471082X13494314

Koenker, R. (2018). *quantreg : Quantile Regression.* R package version 5.36

Koenker, R. et Bache, S. H. (2014). *rqpd : Regression Quantiles for Panel Data.* R package version 0.6/r10

Koenker, R. et Bassett, G. (1982). Tests of linear hypotheses and l_1 estimation. *Econometrica*, *50*(6), 1577–1583.

Koenker, R. et Bassett, Jr., G. (1978). Regression quantiles. *Econometrica. Journal of the Econometric Society*, *46*(1), 33–50.

Koenker, R. et Bilias, Y. (2002). *Quantile regression for duration data : A reappraisal of the Pennsylvania Reemployment Bonus Experiments*, Dans B. Fitzenberger, R. Koenker, et J. A. F. Machado (dir.). *Economic Applications of Quantile Regression*, (p. 199–220). Physica-Verlag HD : Heidelberg

Koenker, R. et Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, *94*(448), 1296–1310.

Koenker, R. et Vasco, D. (1987). Algorithm as 229 : Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *36*(3), 383–393.

Koop, G. et Tobias, J. L. (2004). Learning about heterogeneity in returns to schooling. *Journal of Applied Econometrics*, *19*(7), 827–849.

Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, *157*(2), 396–408. http://dx.doi.org/10.1016/j.jeconom.2010.03.042

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, *95*(2), 391 – 413. http://dx.doi.org/https://doi.org/10.1016/S0304-4076(99)00044-5

Lange, K. (2013). *Optimization*. Springer Texts in Statistics. Springer.

Lee, Y. et Nelder, J. A. (2004). Conditional and marginal models : Another view. *Statistical Science*, *19*(2), 219–228.

Lemieux, T. (2006). *The "Mincer Equation" Thirty Years After Schooling, Experience, and Earnings*, Dans S. Grossbard (dir.). *Jacob Mincer A Pioneer of Modern Labor Economics*, (p. 127–145). Springer US : Boston, MA

Leng, C. et Zhang, W. (2014). Smoothing combined estimating equations in quantile regression for longitudinal data. *Statistics and Computing*, *24*(1), 123–136.

Li, C., Wei, Y., Chappell, R. et He, X. (2011). Bent line quantile regression with application to an allometric study of land mammals' speed and mass. *Biometrics*, *67*(1), 242–249. http://dx.doi.org/10.1111/j.1541-0420.2010.01436.x

Liang, K.-Y. et Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Lid Hjort, N. et Pollard, D. (2011). Asymptotics for minimisers of convex processes. *ArXiv e-prints*.

Liu, Y. et Wu, Y. (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *Journal of Nonparametric Statistics*, *23*(2), 415–437.

Lovering, J. (1888). The "mécanique céleste" of laplace, and its translation, with a commentary by bowditch. *Proceedings of the American Academy of Arts and Sciences*, *24*, 185–201.

Lu, X. et Fan, Z. (2015). Weighted quantile regression for longitudinal data. *Computational Statistics*, *30*(2), 569–592. http://dx.doi.org/10.1007/s00180-014-0550-x

Machado, J. A. F. et Silva, J. M. C. S. (2005). Quantiles for counts. *Journal of the American Statistical Association*, *100*(472), 1226–1237.

Majumdar, A. et Paul, D. (2016). Zero expectile processes and bayesian spatial regression. *Journal of Computational and Graphical Statistics*, *25*(3), 727–747. http://dx.doi.org/10.1080/10618600.2015.1062014

McCullagh, P. et Nelder, J. (1989). *Generalized Linear Models, Second Edition.* Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. Chapman & Hall.

Mersmann, O. (2018). *microbenchmark : Accurate Timing Functions.* R package version 1.4-6

Mkhadri, A., Ouhourane, M. et Oualkacha, K. (2017). A coordinate descent algorithm for computing penalized smooth quantile regression. *Statistics and Computing*, *27*(4), 865–883. http://dx.doi.org/10.1007/s11222-016-9659-9

Newey, W. K. et Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, *55*(4), 819–47.

Neyman, J. et Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, *16*(1), 1–32.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics*, *57*(1), 120–125.

Pollard, D. (1991a). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, *7*(2), 186–199.

Pollard, D. (1991b). Asymptotics for Least Absolute Deviation Regression Estimators. *Econometric Theory*, *7*(2), 186–199.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, *32*(1), 143–155.

R Core Team (2017). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria

R Core Team (2018). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabe-Hesketh, S. et Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata, Second Edition* (2nd éd.). Stata Press.

Righi, M., Yang, Y. et Ceretta, P. (2014). Nonparametric expectile regression for conditional autoregressive expected shortfall estimation. *Contemporary Studies in Economic and Financial Analysis*, *96*, 83–95. http://dx.doi.org/10.1108/S1569-375920140000096003

Robinson, G. K. (1991). That blup is a good thing : The estimation of random effects. *Statistical Science*, *6*(1), 15–32.

Rossi, G. D. et Harvey, A. (2009). Quantiles, expectiles and splines. *Journal of Econometrics*, *152*(2), 179 – 185. http://dx.doi.org/http://dx.doi.org/10.1016/j.jeconom.2009.01.001

Rothman, K., Greenland, S. et Lash, T. (2008). *Modern Epidemiology*. Wolters Kluwer Health/Lippincott Williams & Wilkins.

Ruppert, D. et Carroll, R. J. (1980). Trimmed Least Squares Estimation in the Linear Model. *Journal of the American Statistical Association*, *75*(372), 828–838. http://dx.doi.org/10.2307/2287169

Schempf, A. H. et Kaufman, J. S. (2012). Accounting for context in studies of health inequalities : a review and comparison of analytic approaches. *Annals of Epidemiology*, *22*(10), 683 – 690. http://dx.doi.org/https://doi.org/10.1016/j.annepidem.2012.06.105. Récupéré de http://www.sciencedirect.com/science/article/pii/S1047279712002955

Schnabel, S. et Eilers, P. (2009). Optimal expectile smoothing. *Computational Statistics and Data Analysis*, *53*(12), 4168–4177. http://dx.doi.org/10.1016/j.csda.2009.05.002

Schnabel, S. et Eilers, P. (2013). A location-scale model for non-crossing expectile curves. *Stat*, *2*(1), 171–183. http://dx.doi.org/10.1002/sta4.27

Schulze, W. L. et Kauermann, G. (2015). Smooth expectiles for panel data using penalized splines. *Statistics and Computing*, *27*(1), 271–282. http://dx.doi.org/10.1007/s11222-015-9621-2

Smith, L., Fuentes, M., Gordon-Larsen, P. et Reich, B. (2015). Quantile regression for mixed models with an application to examine blood pressure trends in china. *Annals of Applied Statistics*, *9*(3), 1226–1246.

Sobotka, F., Kauermann, G., Waltrup, L. S. et Kneib, T. (2013). On confidence intervals for semiparametric expectile regression. *Statistics and Computing*, *23*(2), 135–148. http://dx.doi.org/10.1007/s11222-011-9297-1

Sobotka, F. et Kneib, T. (2012). Geoadditive expectile regression. *Computational Statistics and Data Analysis*, *56*(4), 755–767. http://dx.doi.org/10.1016/j.csda.2010.11.015

Stefanie, M., Leonhard, H. et F., K. L. (2016). Marginal or conditional regression models for correlated non-normal data? *Methods in Ecology and Evolution*, *7*(12), 1514–1524. http://dx.doi.org/10.1111/2041-210X.12623

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288.

Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, *109*(3), 475–494. http://dx.doi.org/10.1023/A:1017501703105

Waldmann, E., Sobotka, F. et Kneib, T. (2016). Bayesian regularisation in geoadditive expectile regression. *Statistics and Computing*, 1–15. http://dx.doi.org/10.1007/s11222-016-9703-9

Waltrup, L., Sobotka, F., Kneib, T. et Kauermann, G. (2015a). Expectile and quantile regression—David and Goliath? *Statistical Modelling*, *15*(5), 433–456. http://dx.doi.org/10.1177/1471082X14561155

Waltrup, L. S., Sobotka, F., Kneib, T. et Kauermann, G. (2015b). Expectile and quantile regression—david and goliath ? *Statistical Modelling*.

Wang, H. et Leng, C. (2007). Unified lasso estimation by least squares approximation. *Journal of the American Statistical Association*, *102*(479), 1039–1048.

Wang, H., Li, G. et Jiang, G. (2007a). Robust regression shrinkage and consistent variable selection through the lad-lasso. *Journal of Business & Economic Statistics*, *25*(3), 347–355.

Wang, H., Li, R. et Tsai, C.-L. (2007b). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, *94*(3), 553–568.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–38.

White, H. (2001). *Asymptotic Theory for Econometricians*. Economic theory, econometrics, and mathematical economics. Academic Press.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*, volume 1 de *MIT Press Books*. The MIT Press.

Wu, T. T. et Lange, K. (2008). Coordinate descent algorithms for lasso penalized regression. *Ann. Appl. Stat.*, *2*(1), 224–244. http://dx.doi.org/10.1214/07-AOAS147

Xing, J.-J. et Qian, X.-Y. (2017). Bayesian expectile regression with asymmetric normal distribution. *Communications in Statistics - Theory and Methods*, *46*(9), 4545–4555. http://dx.doi.org/10.1080/03610926.2015.1088030

Xu, Q., Liu, X., Jiang, C. et Yu, K. (2016). Nonparametric conditional autoregressive expectile model via neural network with applications to estimating financial risk. *Applied Stochastic Models in Business and Industry*, *32*(6), 882–908. http://dx.doi.org/10.1002/asmb.2212

Yang, Y. et Zou, H. (2015). Nonparametric multiple expectile regression via erboost. *Journal of Statistical Computation and Simulation*, *85*(7), 1442–1458. http://dx.doi.org/10.1080/00949655.2013.876024

Yao, Q. et Tong, H. (1996). Asymmetric least squares regression estimation : A nonparametric approach. *Journal of Nonparametric Statistics*, *6*(2), 273–292. http://dx.doi.org/10.1080/10485259608832675

Yin, G. et Cai, J. (2005). Quantile regression models with multivariate failure time data. *Biometrics*, *61*(1), 151–161. http://dx.doi.org/10.1111/j.0006-341X.2005.030815.x