Weighted asymmetric least squares regression for longitudinal data using GEE

Amadou Barry\textsuperscript{1}, Karim Ouakacha\textsuperscript{1}, Arthur Charpentier\textsuperscript{2}

\textsuperscript{1}Department of Mathematics, Université du Québec à Montréal, Canada,\textsuperscript{2}Faculty of Economics, Université de Rennes 1, France

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-square regression (expectile) to analyze the longitudinal data. Expectile regression naturally extends the classical least squares method and has properties similar to quantile regression. 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, derive their asymptotic properties and 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, working correlation, cluster data, longitudinal data.

1 Introduction

Longitudinal and clustered data arise in many application fields such as epidemiology (Smith et al. 2015), genetics (Furlotte, Eskin, and Eyheramendy 2012), economics (Hsiao 2007), and other fields of biological and social sciences. They are characterized by the presence of a within-subject dependence which 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 high efficient estimators. Generalized estimating equations (GEE) (Liang and 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 and 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 affect 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.

*Corresponding author: barryhafia@gmail.com
ER was introduced by Aigner, Amemiya, and Poirier (1976) under a likelihood-based approach. A decade later, Newey and 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 paper (Efron 1991), expectile lived quietly in the shadows for decades, as mentioned in (Eilers 2013). But recently it has comeback 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 and Eilers 2009, Rossi and Harvey (2009), Sobotka and Kneib (2012), Sobotka et al. (2013)). Other works focused on contrasting ER and 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 and Eilers (2013), Waltrup et al. (2015)). Today ER is extended to many classes of models, such as Bayesian (Majumdar and Paul 2016, Waldmann, Sobotka, and Kneib (2016), Xing and Qian (2017)), nonparametric (Righi, Yang, and Ceretta 2014, Yang and Zou (2015)), nonlinear (Kim and Lee 2016), neural network (Xu et al. 2016, Jiang et al. (2017)), and support vector machine (Farooq and Steinwart 2017). Recently, Waltrup and Kauermann (Schulze Waltrup and 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 and 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, Wei, and Parzen (2004) derived a QR estimator for correlated data using GEE approach based on independence working correlation. Along the same lines, Fu and 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 and Fan (2015) proposed a general stationary auto-correlation 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 (2013b), Waltrup et al. (2015)).

This paper makes its contribution by introducing a new class of estimators for the analysis of dependent data. Section 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 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 4. The GEEE estimator is applied to a real data set and the results are presented in Section 5. the conclusion is presented in Section 6. All proofs are in the appendix.

2 Models and Methods

This Section introduces the univariate expectile and the ER model.

2.1 Expectile and ER model

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

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

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

$$\rho_\tau(t) = |\tau - 1(t \leq 0)| \cdot t^2$$
is the asymmetric square loss function that assigns weights $\tau$ and $1 - \tau$ to positive and negative deviations, respectively.

By equating the first derivative of $\mu_\tau$ to zero, the expectile can also be defined as solution of

$$
\mu_\tau(Y) = \mu = \mu - \frac{1 - 2\tau}{1 - \tau} \mathbb{E} \{ (Y - \mu_\tau(Y)) \mathbb{I}\{Y > \mu_\tau(Y)\} \},
$$

(2)

where $\mu = \mu_{0.5}(Y) = \mathbb{E}(Y)$. This definition, presented by Newey and 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

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

where $\psi_\tau(t) = |\tau - 1(t \leq 0)|$ 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 $\tau$-th empirical expectile

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

is the solution which minimizes the empirical loss function

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

(3)

Newey and 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

$$
y_i = x_i^T \beta + \varepsilon_i,
$$

(4)

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

$$
\mu_\tau(y_i | x_i) = x_i^T \beta_\tau, \quad \text{with} \quad \mu_\tau(\varepsilon_i) = 0.
$$

(5)

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

$$
\frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - x_i^T \beta_\tau)
$$

(6)

over $\beta_\tau \in \mathbb{R}^p$. The asymmetric loss function associated with the expectile function is continuously differentiable, and solving equation (6) gives
\[ \hat{\beta}_\tau = \left( \sum_{i=1}^{n} x_i^T \psi(\tilde{\varepsilon}_{it}) x_i \right)^{-1} \left( \sum_{i=1}^{n} x_i \psi(\tilde{\varepsilon}_{it}) y_i \right), \]

where \( \tilde{\varepsilon}_{it} = y_i - x_i^T \hat{\beta}_\tau \). 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 and Powell (1987) proposed a robust estimator of the variance-covariance matrix of \( \hat{\beta}_\tau \).

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

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 GEEE for longitudinal data

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

\[ y_{it} = x_{it}^T \beta + \varepsilon_{it}, \]

where \( y_{it} \) is the \( t \)-th observation of the response variable for the \( i \)-th individual, \( x_{it} = (x_{1it}, \ldots, x_{pit}) \) is the \( p \times 1 \) covariates, \( \varepsilon_{it} \) the random error and \( \beta \) the \( p \times 1 \) true parameter vector that needs to be estimated.

Equation model \( (9) \) can be conveniently represented in individual notation as

\[ y_i = X_i \beta + \varepsilon_i, \]

where \( y_i \) is the dependent observations of the response variable of the individual \( i \), \( X_i \) the corresponding \( m_i \times p \) matrix covariates and \( \varepsilon_i \) the vector error. Individual observations can also be stacked and presented in matrix form as

\[ y = X \beta + \varepsilon, \]

where \( y \) and \( \varepsilon \) are \( N \times 1 \) vectors, \( X \) 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 \( \tau \) of the model equation \( (9) \) is defined as

\[ \mu_\tau(y_{it} | x_{it}) = x_{it}^T \beta_\tau, \quad \mu_\tau(\varepsilon_{it}) = 0. \]
The assumption, \( \mu_\tau(\varepsilon_{it}) = 0 \), is introduced to guarantee that the random error is centered on the \( \tau \)-th expectile. The parameter \( \beta_\tau \) measures the effect of the covariates \( x_{it} \) 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

\[
S_i(\beta_\tau) = \sum_{i=1}^{n} X_i^T \Psi_\tau(y_i - X_i \beta_\tau) \left[ y_i - X_i \beta_\tau \right] = 0, \tag{13}
\]

where \( \Psi_\tau(y_i - X_i \beta_\tau) = \text{diag} \left( \psi_\tau(y_{i1} - x_{i1}^T \beta_\tau), \ldots, \psi_\tau(y_{im_i} - x_{im_i}^T \beta_\tau) \right) \). The resulting estimator \( \hat{\beta}_{1\tau} \) can also be derived as the minimizer of the following objective function

\[
\frac{1}{N} \sum_{i=1}^{n} \sum_{t=1}^{m_i} \rho_\tau \left( y_{it} - x_{it}^T \beta_\tau \right) \tag{14}
\]

over \( \beta_\tau \in \mathbb{R}^p \). The explicit form of the resulting estimator \( \hat{\beta}_{1\tau} \) is similar to (7).

When \( \tau = 0.5 \), the estimator \( \hat{\beta}_{1\tau} \) corresponds to the GEE estimator introduced by Liang and 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 \( \tau \), the GEEE estimator \( \hat{\beta}_\tau \) is derived by solving the following GEEE equations

\[
S(\beta_\tau) = \sum_{i=1}^{n} X_i^T V_i^{-1} \Psi_\tau(y_i - X_i \beta_\tau) \left[ y_i - X_i \beta_\tau \right] = 0, \tag{15}
\]

where \( V_i \) is a working covariance matrix represented as

\[
V_i = \sigma_\tau^2 A_i^2 R_i(\alpha_\tau) A_i^{\frac{3}{2}}, \tag{16}
\]

and \( \sigma_\tau^2 \) is the nuisance parameter. \( A_i \) is the \( m_i \times m_i \) diagonal matrix with the variance function \( \nu(\mu_i) \) as diagonal elements and \( R_i(\alpha_\tau) \) as the working correlation matrix.

The working correlation matrix \( R_i(\alpha_\tau) \) describes the correlation pattern of within-subject observations with the \( K \times 1 \) vector parameter \( \alpha_\tau \). Liang and 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 (13). The GEEE exchangeable structure is a simple extension of the independence working correlation. It assumes a common correlation, \( \rho_{ts} = \alpha_\tau, \forall t \neq s \), 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, \( \rho_{ts} = \alpha_\tau |t-s| \). This structure assigns the highest correlation to adjacent pairs of observations 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, \( \rho_{ts} = \alpha_{ts} \), \( \forall t \neq s \).

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 \( \tilde{\beta}_1^{(0)} = \tilde{\beta}_1 \), the estimator obtained from (13).

**Step 2.** Given \( \tilde{\beta}_1^{(r-1)} \) from the \( r-1 \) step update

\[
\tilde{\sigma}_1^{2(r)} \leftarrow \frac{1}{N-p} \sum_{i=1}^{n} \sum_{t=1}^{m_i} \psi_1(\tilde{e}_{itt})^2 \tilde{e}_{itt}^2,
\]

\[
\tilde{\alpha}_1^{(r)} \leftarrow \frac{1}{(N-1)\tilde{\sigma}_1^{2(r)}} \sum_{i=1}^{n} \sum_{t<s} \psi_1(\tilde{e}_{itt}) \tilde{e}_{itt} \psi_1(\tilde{e}_{ist}) \tilde{e}_{ist},
\]

where \( N_1 = \frac{1}{2} \sum_{i=1}^{n} m_i(m_i - 1) \) and \( \tilde{e}_{itt} = y_{it} - x_{it}^T \tilde{\beta}_1^{(r-1)} \).

**Step 3.** Update \( \tilde{\beta}_1^{(r)} \) by

\[
\tilde{\beta}_1^{(r)} \leftarrow \tilde{\beta}_1^{(r-1)} + \left[ \sum_{i=1}^{n} X_i^TV_i^{-1}(\tilde{\alpha}_1^{(r-1)})\Psi_1(\tilde{\beta}_1^{(r-1)})X_i \right]^{-1} S(\tilde{\alpha}_1^{(r-1)}, \tilde{\beta}_1^{(r-1)}),
\]

where \( \Psi_1(\tilde{\beta}_1^{(r-1)}) = \Psi_1(y_i - X_i\tilde{\beta}_1^{(r-1)}) \).

**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 \( \alpha \) 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 \( \alpha_1 \) is estimated by

\[
\tilde{\alpha}_1 = \frac{1}{(N_2 - p)\tilde{\sigma}_1^{2(r)}} \sum_{i=1}^{n} \sum_{t<s} \psi_1(\tilde{e}_{itt}) \tilde{e}_{itt} \psi_1(\tilde{e}_{i,t+1}, \tilde{e}_{it+1}, \tilde{e}_{ist}), \quad N_2 = \sum_{i=1}^{n} (m_i - 1).
\]

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

\[
\tilde{\alpha}_{1st} = \frac{1}{(N-p)\tilde{\sigma}_1^{2(r)}} \sum_{i=1}^{n} \psi_1(\tilde{e}_{itt}) \tilde{e}_{itt} \psi_1(\tilde{e}_{ist}, \tilde{e}_{ist}).
\]

Generalization to other GEEE-working correlations is straightforward.

In Section 3 it shown that the GEEE estimator \( \tilde{\beta}_1 \) is consistent and asymptotically normally distributed.
In addition, the simulation results of Section 4 show that the GEEE method yields a consistent and highly
efficient estimator even with a misspecification of the true covariance structure.
2.3 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 and Wu 2011). For a fixed \( \tau = (\tau_1, \ldots, \tau_q) \) the GEEE estimating functions are defined as

\[
S(\beta_\tau) = \sum_{k=1}^{q} S_{\tau_k}(\beta_{\tau_k}) = \sum_{i=1}^{n} (W \otimes X_i)^T V_{\tau}^{-1} \Psi_{\tau} \left( I_q \otimes y_i - (I_q \otimes X_i) \beta_{\tau} \right),
\]

where \( S_{\tau_k} \) is defined in (15), and \( W = \text{diag}(w_k)_{k=1}^{q} \) is the \( q \times q \) matrix of weights controlling the relative influence of the \( q \) expectiles. \( V_{\tau} \) is given by (16) and

\[
\Psi_{\tau} \left( I_q \otimes y_i - (I_q \otimes X_i) \beta_{\tau} \right) = \text{diag} \left( \Psi_{\tau_1}(y_i - X_i \beta_{\tau_1}), \ldots, \Psi_{\tau_q}(y_i - X_i \beta_{\tau_q}) \right).
\]

The parameter \( \beta_\tau \) is obtained using the iterative re-weighted least squares algorithm as shown above for a single expectile.

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

3 Asymptotic properties

This Section presents the asymptotic properties of the GEEE estimator for several fixed expectiles \( \tau \). In the first step, the asymptotic properties of the GEEE estimator \( \hat{\beta}_{f,\tau} \) with the independent working correlation structure are presented. Subsequently, the asymptotic properties of the GEEE estimator \( \hat{\beta}_\tau \) with a general correlation structure are derived. The main reason for presenting the results of \( \hat{\beta}_{f,\tau} \) separately is that; it is also the estimator of the expectile regression for a marginal model based on the AND distribution (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.

3.1 Asymptotic properties for the independent GEEE

To begin, assume the following conditions.

A1. The data \( \{(y_i, X_i)\}_{i=1}^{n} \) are independent across \( i \), and

\[
\text{Var} \left[ \Psi_{\tau}(\epsilon_{i,\tau}) \epsilon_{i,\tau} \right] = E \left[ \Psi_{\tau}(\epsilon_{i,\tau}) \epsilon_{i,\tau} \epsilon_{i,\tau} \Psi_{\tau}(\epsilon_{i,\tau}) \right] = \Sigma_{i,\tau}, \text{ where } \epsilon_{i,\tau} = \left( \epsilon_{i,\tau_1}, \ldots, \epsilon_{i,\tau_q} \right)^T, \epsilon_{i,\tau_k} = (e_{i1,\tau_k}, \ldots, e_{im_i,\tau_k})^T, \epsilon_{i,\tau_k} = y_{it} - x_{it}^T \beta_{\tau_k} \text{ and } \Psi_{\tau}(\epsilon_{i,\tau}) = \left[ \text{diag}(\Psi_{\tau_k}(\epsilon_{i,\tau_k})) \right]_{k=1}^{q}.
\]

A2. The limiting forms of the following matrices are positive definite.
\[
D_{11}(\tau) = \lim_{n \to \infty} N^{-1} \sum_{i=1}^{n} (W \otimes X_i)^T \Psi_\tau(\varepsilon_i \mid \beta)(I_q \otimes X_i),
\]

\[
D_{10}(\tau) = \lim_{n \to \infty} N^{-1} \sum_{i=1}^{n} (W \otimes X_i)^T \Sigma_\tau(W \otimes X_i).
\]

**A3.** max_{1 \leq i \leq n, 1 \leq i \leq m,} \|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 \(\Sigma_{i0.5} = 1/4 \text{Var}[\varepsilon_{i0.5}]\) becomes the variance of \(\varepsilon_i\) up to a factor and \(D_{10} = 1/4 \lim_{n \to \infty} N^{-1} \sum_{i=1}^{n} X_i^T \text{Var}[\varepsilon_{i0.5}]X_i\). Considering \(D_{11} = 1/2 \lim_{n \to \infty} N^{-1} \sum_{i=1}^{n} X_i^T X_i\), we see that this factor disappears in the expression of the variance of the estimator. Hence, when \(\tau = 1/2\), the condition A2 is reduced to a condition on the matrices \(N^{-1} \sum_{i=1}^{n} X_i^T \text{Var}[\varepsilon_i]X_i\) and \(N^{-1} \sum_{i=1}^{n} X_i^T X_i\). Condition A3 is important both for the convergence and for the Lindeberg condition. The following **Theorem 1** states the results of the asymptotic properties of the GEE estimator \(\hat{\beta}_{i\tau}\) assuming an independent working correlation structure.

**Theorem 1.** Assume that \(\hat{\beta}_{i\tau}\) is the solution of the estimating function (13) and suppose the data are generated by model (9), and that conditions A1-A3 are satisfied. If \(\text{E}[\psi_{i\tau}\varepsilon_{i\tau}^{\tau}]|^{4+\nu} < \Delta \) and \(\text{E}[\varepsilon_{i\tau}^{\tau}]|^{4+\nu} < \Delta\) for some \(\nu > 0\) and \(\Delta > 0\), then for every fixed sequence of expectiles \(\tau = (\tau_1, \ldots, \tau_q)\)

\[
\sqrt{N}(\hat{\beta}_{i\tau} - \beta_{\tau}) \overset{d}{\to} \mathcal{N}(0, D_{11}^{-1}(\tau)D_{10}(\tau)D_{11}^{-1}(\tau)).
\]

In order to use this new estimator \(\hat{\beta}_{i\tau}\) to make inference, an estimator of its VC-matrix is presented in **Theorem 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 and 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,

\[
\hat{D}_{11}(\tau) = N^{-1} \sum_{i=1}^{n} (W \otimes X_i)^T \hat{\Psi}_\tau(\hat{\varepsilon}_i \mid \hat{\beta})(I_q \otimes X_i),
\]

\[
\hat{D}_{10}(\tau) = N^{-1} \sum_{i=1}^{n} (W \otimes X_i)^T \hat{\Sigma}_\tau(W \otimes X_i)
\]

where \(\hat{\Sigma}_\tau = \Psi_\tau(\hat{\varepsilon}_i \mid \hat{\beta} \hat{\varepsilon}_i \hat{\beta}^T \Psi_\tau(\hat{\varepsilon}_i \mid \hat{\beta})\) and \(\hat{\varepsilon}_i \) is obtained by replacing \(\beta_{\tau}\) with \(\hat{\beta}_{i\tau}\) in the expression of \(\varepsilon_{i\tau}\). Then, we have **Theorem 2**

**Theorem 2.** Suppose the data are generated by model (9) and that conditions A1-A3 are satisfied. If \(\text{E}[\psi_{i\tau}\varepsilon_{i\tau}^{\tau}]|^{4+\nu} < \Delta \) and \(\text{E}[\varepsilon_{i\tau}^{\tau}]|^{4+\nu} < \Delta\) for some \(\nu > 0\) and \(\Delta > 0\), then for every fixed sequence of expectiles \(\tau = (\tau_1, \ldots, \tau_q)\)

\[
\hat{D}_{11}^{-1}(\tau)\hat{D}_{10}(\tau)\hat{D}_{11}^{-1}(\tau) \overset{p}{\to} D_{11}^{-1}(\tau)D_{10}(\tau)D_{11}^{-1}(\tau).
\]
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 \( \{(y_i, X_i)\}_{i=1}^{m} \) are independent across \( i \) and \( \text{Var} \left[ \Psi_{i\tau}(\varepsilon_{i\tau})\varepsilon_{i\tau} \right] = \Sigma_{i\tau} \).

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

\[
D_1(\tau) = \lim_{n \to \infty} N^{-1} \sum_{i=1}^{n} (W \otimes X_i)^T V_{i\tau}^{-1} E[\Psi_{i\tau}(\varepsilon_{i\tau})](I_i \otimes X_i),
\]

\[
D_0(\tau) = \lim_{n \to \infty} N^{-1} \sum_{i=1}^{n} (W \otimes X_i)^T V_{i\tau}^{-1} \Sigma_{i\tau} V_{i\tau}^{-1}(W \otimes X_i).
\]

B3. \( \max_{1 \leq i \leq n} \|x_{it}\| < M. \)

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

Theorem 3. Suppose the data are generated by model (7) and that conditions B1-B3 are satisfied. If \( E[\Psi_{i\tau}(\varepsilon_{i\tau})]^{4+\nu} < \Delta \) and \( E|\varepsilon_{i\tau}|^{4+\nu} < \Delta \) for some \( \nu > 0 \) and \( \Delta > 0 \), then for every fixed sequence of expectiles \( \tau = (\tau_1, \ldots, \tau_q) \)

\[
\sqrt{N}(\hat{\beta}_\tau - \beta_\tau) \overset{d}{\to} N\left(0, D_1^{-1}(\tau)D_0(\tau)D_1^{-1}(\tau)\right).
\]

In the same way as with the GEEE-independent working correlation estimator, the following Theorem 4 proposes an estimator of the VC-matrix of estimator \( \hat{\beta}_\tau \). Consider \( \hat{V}_{i\tau} \) to be a consistent estimator of \( V_{i\tau} \).

Then, under the above conditions, Theorem 4 is stated as follows

Theorem 4. Suppose the data are generated by model (7) and that conditions B1-B3 are satisfied. Assume \( E[\Psi_{i\tau}(\varepsilon_{i\tau})]^{4+\nu} < \Delta \) and \( E|\varepsilon_{i\tau}|^{4+\nu} < \Delta \) for some \( \nu > 0 \) and \( \Delta > 0 \). Then for every fixed sequence of expectiles \( \tau = (\tau_1, \ldots, \tau_q) \)

\[
\hat{D}_1^{-1}(\tau)\hat{D}_0(\tau)\hat{D}_1^{-1}(\tau) \overset{d}{\to} D_1^{-1}(\tau)D_0(\tau)D_1^{-1}(\tau)
\]

where

\[
\hat{D}_1(\tau) = N^{-1} \sum_{i=1}^{n} (W \otimes X_i)^T \hat{V}_{i\tau}^{-1} \Psi_{i\tau}(\hat{\varepsilon}_{i\tau})(I_i \otimes X_i),
\]

\[
\hat{D}_0(\tau) = N^{-1} \sum_{i=1}^{n} (W \otimes X_i)^T \hat{V}_{i\tau}^{-1} \hat{\Sigma}_{i\tau} \hat{V}_{i\tau}^{-1}(W \otimes X_i),
\]

and \( \hat{\Sigma}_{i\tau} = \Psi_{i\tau}(\hat{\varepsilon}_{i\tau})\hat{\varepsilon}_{i\tau}\hat{\varepsilon}_{i\tau}^T \Psi_{i\tau}(\hat{\varepsilon}_{i\tau}). \)

4 Simulation

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 \( (M_\gamma) \) :
\[ y_{it} = \beta_0 + x_{it}\beta_1 + (1 + \gamma x_{it})\varepsilon_{it}, \ i = 1, \ldots, n \text{ and } t = 1, \ldots, m_i. \] (18)

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

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 \( \beta_0 \) and \( \beta_1 \) 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 \( \tau \)-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: \( \varepsilon_{it} = \varepsilon_{it}' - \mu_T(\varepsilon_{it}') \);
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 integer 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 \texttt{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) of the estimates are reported for the measurement of the quality of the different related estimators. The standard deviation (SD) of the 400 parameter estimates is used as a benchmark to evaluate the average asymptotic standard errors (SE).

We use the quasi-likelihood criterion (QIC) as a model-selection criteria to choose among the different working correlation structures. The QIC is a criteria 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

\[
QIC(R) = \frac{1}{2} \sum_{i=1}^{n} \sum_{t=1}^{m_i} \frac{\varepsilon_{it}^2}{\hat{\sigma}^2} + 2 \text{Trace} \left( \hat{\Omega}_i \hat{V}_R(\hat{\beta}) \right),
\] (19)

where \( \hat{V}_R(\hat{\beta}) \) is the robust covariance estimate and \( \hat{\Omega}_i \) is the inverse of the covariance estimate under the independent working correlation evaluated at \( \hat{\beta}_T(R) \), 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 and Bottai 2007, Geraci and Bottai (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 and 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 1 and Table 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 $\rho$. The Un and AR1 estimators do slightly better than the Ind estimator in term of efficiency, particularly when the correlation is higher $\rho \in (0.5, 0.9)$. The Un estimator do much better in general than the other three estimators (Ind, Exc, AR1).

To evaluate the asymptotic standard error (SE), we use the standard deviation (SD) as a benchmark. The results are presented in Table 3 and Table 4 when the error follows a normal distribution respectively for the $(M_0)$ and $(M_{1/10})$ models. We observe that the values of SD and SE decrease as $n$ becomes large. In general, the values of SD and SE are identical for each of the estimators. This identity is more pronounced in the case of the independent and unstructured working correlation. 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. Hence 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 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 is not surprising. Similar results have been reported in (Jang 2011).

As in Pan’s paper (Jang 2011), many 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 further details.

The last Tables 6, 7 report the simulation results (Bias and RMSE) of the lqmm estimator and the GEEE estimator with exchangeable working correlation. The results show that both methods are competitive in term of Bias and RMSE.
Table 1: Bias and relative efficiency of GEEE estimator with different correlation structures at 3 percentiles with $\rho \in (0.1, 0.5, 0.9)$, and $\varepsilon \sim \mathcal{N}(0, 1)$ under a location-shift scenario.

| $\tau$ | $\rho$ | $m = 4$ |  |  |  | $m \sim \mathcal{U}(3, 7)$ |  |  |  |
|---|---|---|---|---|---|---|---|---|---|
|   |   |   | 50 | 100 |   | 50 | 100 |   |   |
| 0.25 | 0.1 | Ind | -0.0006 | 1.000 | -0.0003 | 1.000 | 0.0006 | 1.000 | -0.0007 | 1.000 |
|     |     | AR1 | -0.0006 | 1.000 | -0.0003 | 1.010 | 0.0007 | 1.000 | -0.0006 | 1.000 |
|     |     | Un  | -0.0003 | 1.170 | -0.0001 | 1.099 | 0.0006 | 1.254 | -0.0006 | 1.167 |
| 0.5  |     |     | -0.0007 | 1.000 | 0.0000 | 1.000 | 0.0001 | 1.000 | -0.0010 | 1.000 |
|     |     |     | AR1  | -0.0007 | 1.000 | 0.0000 | 1.000 | 0.0005 | 1.000 | -0.0008 | 1.000 |
|     |     | Un  | -0.0007 | 1.142 | 0.0001 | 1.083 | 0.0002 | 1.233 | -0.0008 | 1.140 |
| 0.9  |     |     | -0.0006 | 1.000 | 0.0000 | 1.000 | 0.0001 | 1.000 | -0.0010 | 1.000 |
|     |     |     | AR1  | -0.0006 | 1.007 | 0.0000 | 1.000 | 0.0005 | 1.000 | -0.0008 | 1.000 |
|     |     | Un  | -0.0007 | 1.164 | 0.0002 | 1.089 | 0.0002 | 1.272 | -0.0010 | 1.167 |
| 0.50 | 0.1 | Ind | 0.0001 | 1.000 | -0.0007 | 1.000 | -0.0002 | 1.000 | 0.0002 | 1.000 |
|     |     | AR1 | 0.0001 | 1.079 | -0.0007 | 1.070 | -0.0002 | 1.081 | 0.0002 | 1.089 |
|     |     | Un  | -0.0003 | 1.664 | -0.0002 | 1.600 | -0.0003 | 1.790 | 0.0004 | 1.722 |
| 0.5  |     |     | 0.0000 | 1.000 | -0.0005 | 1.000 | -0.0003 | 1.000 | 0.0002 | 1.000 |
|     |     |     | AR1  | 0.0001 | 1.104 | -0.0005 | 1.105 | -0.0003 | 1.118 | 0.0002 | 1.116 |
|     |     | Un  | -0.0002 | 1.687 | -0.0001 | 1.653 | -0.0004 | 1.807 | 0.0003 | 1.733 |
| 0.9  |     |     | 0.0000 | 1.000 | -0.0003 | 1.000 | -0.0003 | 1.000 | 0.0001 | 1.000 |
|     |     |     | AR1  | 0.0001 | 1.080 | -0.0003 | 1.070 | -0.0004 | 1.080 | 0.0001 | 1.067 |
|     |     | Un  | -0.0003 | 1.681 | 0.0000 | 1.590 | -0.0007 | 1.784 | 0.0001 | 1.678 |
| 0.75 | 0.1 | Ind | -0.0009 | 1.000 | 0.0003 | 1.000 | 0.0003 | 1.000 | 0.0001 | 1.000 |
|     |     | AR1 | -0.0005 | 1.277 | 0.0002 | 1.265 | 0.0004 | 1.294 | 0.0000 | 1.284 |
|     |     | Un  | -0.0007 | 3.328 | -0.0001 | 3.316 | 0.0001 | 3.698 | 0.0005 | 3.682 |
| 0.5  |     |     | -0.0009 | 1.000 | 0.0004 | 1.000 | 0.0005 | 1.000 | 0.0001 | 1.000 |
|     |     |     | AR1  | -0.0005 | 1.376 | 0.0003 | 1.362 | 0.0005 | 1.408 | 0.0000 | 1.405 |
|     |     | Un  | -0.0001 | 3.391 | 0.0001 | 3.351 | 0.0001 | 3.783 | 0.0002 | 3.762 |
| 0.9  |     |     | -0.0008 | 1.000 | 0.0006 | 1.000 | 0.0006 | 1.000 | 0.0001 | 1.000 |
|     |     |     | AR1  | -0.0006 | 1.273 | 0.0005 | 1.265 | 0.0005 | 1.298 | 0.0000 | 1.284 |
|     |     | Un  | -0.0007 | 3.367 | 0.0000 | 3.306 | 0.0005 | 3.823 | -0.0001 | 3.670 |
Table 2: Bias and relative efficiency of GEEE estimator with different correlation structures at 3 percentiles with $\rho \in (0.1, 0.5, 0.9)$, and $\varepsilon \sim \mathcal{N}(0, 1)$ under a location-scale-shift scenario.

| $\tau$ | $\rho$ | GEEE | 50  | 100  | 50  | 100  |
|--------|--------|------|-----|------|-----|------|
|        |        |      | Bias   | EFF | Bias   | EFF |
| 0.25   | 0.1    | Ind  | 0.0013 | 1.000 | 0.0005 | 1.000 | 0.007 | 1.000 | 0.0006 | 1.000 |
|        |        | AR1  | -0.0014 | 1.761 | -0.0078 | 1.611 | 0.0074 | 2.272 | -0.0101 | 1.717 |
|        |        | Un   | 0.0035 | 2.608 | 0.0006 | 2.365 | 0.0048 | 2.693 | -0.0002 | 2.974 |
| 0.5    |        | Ind  | -0.0006 | 1.000 | -0.0001 | 1.000 | 0.0001 | 1.000 | 0.0003 | 1.000 |
|        |        | AR1  | 0.0040 | 1.853 | -0.0003 | 1.484 | 0.0107 | 2.806 | 0.0023 | 1.671 |
|        |        | Un   | 0.0021 | 2.338 | 0.0002 | 2.081 | 0.0017 | 2.352 | -0.0008 | 2.470 |
| 0.9    |        | Ind  | -0.0025 | 1.000 | -0.0009 | 1.000 | -0.0008 | 1.000 | 0.0002 | 1.000 |
|        |        | AR1  | 0.0120 | 1.830 | 0.0083 | 1.739 | 0.0128 | 3.492 | 0.0080 | 17.936 |
|        |        | Un   | 0.0009 | 2.752 | 0.0007 | 2.636 | -0.0009 | 3.020 | -0.0017 | 3.167 |
| 0.50   | 0.1    | Ind  | -0.0022 | 1.000 | 0.0012 | 1.000 | 0.0015 | 1.000 | 0.0003 | 1.000 |
|        |        | AR1  | -0.0039 | 1.758 | -0.0023 | 1.780 | 0.0045 | 3.080 | -0.0005 | 2.646 |
|        |        | Un   | 0.0042 | 2.390 | 0.0021 | 2.304 | -0.0011 | 3.219 | -0.0008 | 2.994 |
| 0.5    |        | Ind  | -0.0031 | 1.000 | 0.0001 | 1.000 | -0.0004 | 1.000 | -0.0001 | 1.000 |
|        |        | AR1  | 0.0044 | 1.678 | 0.0028 | 1.699 | 0.0094 | 3.172 | 0.0054 | 2.766 |
|        |        | Un   | 0.0034 | 2.075 | 0.0017 | 2.030 | 0.0006 | 2.778 | 0.0002 | 2.481 |
| 0.9    |        | Ind  | -0.0040 | 1.000 | -0.0009 | 1.000 | -0.0020 | 1.000 | -0.0006 | 1.000 |
|        |        | AR1  | 0.0128 | 1.813 | 0.0084 | 1.836 | 0.0195 | 3.217 | 0.0131 | 2.726 |
|        |        | Un   | 0.0026 | 2.461 | 0.0025 | 2.591 | 0.0029 | 3.429 | 0.0008 | 3.140 |
| 0.75   | 0.1    | Ind  | 0.0000 | 1.000 | 0.0024 | 1.000 | -0.0010 | 1.000 | 0.0013 | 1.000 |
|        |        | AR1  | 0.0002 | 2.312 | 0.0032 | 2.207 | -0.0165 | 3.860 | -0.0049 | 3.845 |
|        |        | Un   | -0.0009 | 2.192 | 0.0017 | 2.151 | 0.0000 | 2.634 | -0.0019 | 2.805 |
| 0.5    |        | Ind  | -0.0015 | 1.000 | 0.0009 | 1.000 | -0.0023 | 1.000 | 0.0006 | 1.000 |
|        |        | AR1  | -0.0019 | 2.150 | 0.0009 | 2.114 | -0.0197 | 3.792 | -0.0038 | 3.700 |
|        |        | Un   | -0.0029 | 1.850 | -0.0006 | 1.892 | -0.0032 | 2.248 | -0.0022 | 2.159 |
| 0.9    |        | Ind  | -0.0024 | 1.000 | -0.0006 | 1.000 | -0.0039 | 1.000 | -0.0001 | 1.000 |
|        |        | AR1  | -0.0042 | 2.210 | -0.0001 | 2.251 | -0.0235 | 3.785 | -0.0027 | 3.701 |
|        |        | Un   | -0.0079 | 1.931 | -0.0013 | 2.328 | -0.0069 | 2.544 | -0.0044 | 2.494 |
Table 3: Standard deviation and asymptotic standard errors of the GEEE estimator with different correlation structures at 3 percentiles with $\rho \in (0.1, 0.5, 0.9)$, and $\varepsilon \sim \mathcal{N}(0,1)$ under a location-shift scenario.

| $\tau$ | $\rho$ | $m = 4$ |       |       | $m \sim \mathcal{U}(3, 7)$ |       |       |
|-------|-------|---------|-------|-------|----------------------------|-------|-------|
|       |       | 50      | 100   |       | 50                        | 100   |       |
|       |       | SD  | SE | SD  | SE | SD  | SE | SD  | SE |
| 0.25  | 0.1   | 0.014 | 0.014 | 0.010 | 0.010 | 0.014 | 0.013 | 0.009 | 0.009 |
|       |       | 0.014 | 0.014 | 0.010 | 0.010 | 0.014 | 0.013 | 0.009 | 0.009 |
|       |       | 0.014 | 0.016 | 0.010 | 0.011 | 0.014 | 0.016 | 0.009 | 0.010 |
| 0.5   | 0.1   | 0.013 | 0.013 | 0.010 | 0.010 | 0.013 | 0.012 | 0.009 | 0.009 |
|       |       | 0.013 | 0.014 | 0.010 | 0.010 | 0.013 | 0.012 | 0.009 | 0.009 |
|       |       | 0.014 | 0.015 | 0.010 | 0.010 | 0.013 | 0.015 | 0.009 | 0.010 |
| 0.9   | 0.1   | 0.014 | 0.014 | 0.010 | 0.010 | 0.014 | 0.012 | 0.009 | 0.009 |
|       |       | 0.014 | 0.014 | 0.010 | 0.010 | 0.014 | 0.012 | 0.009 | 0.009 |
|       |       | 0.015 | 0.016 | 0.010 | 0.011 | 0.014 | 0.016 | 0.009 | 0.010 |
| 0.50  | 0.1   | 0.014 | 0.014 | 0.010 | 0.010 | 0.012 | 0.012 | 0.009 | 0.009 |
|       |       | 0.013 | 0.015 | 0.009 | 0.011 | 0.012 | 0.013 | 0.008 | 0.010 |
|       |       | 0.013 | 0.023 | 0.008 | 0.016 | 0.011 | 0.022 | 0.008 | 0.016 |
| 0.5   | 0.5   | 0.014 | 0.013 | 0.010 | 0.010 | 0.012 | 0.012 | 0.009 | 0.009 |
|       |       | 0.012 | 0.015 | 0.009 | 0.010 | 0.011 | 0.013 | 0.008 | 0.010 |
|       |       | 0.012 | 0.023 | 0.008 | 0.016 | 0.010 | 0.022 | 0.007 | 0.015 |
| 0.9   | 0.5   | 0.014 | 0.014 | 0.010 | 0.010 | 0.012 | 0.012 | 0.009 | 0.009 |
|       |       | 0.013 | 0.015 | 0.010 | 0.011 | 0.012 | 0.014 | 0.008 | 0.010 |
|       |       | 0.013 | 0.023 | 0.009 | 0.016 | 0.011 | 0.022 | 0.008 | 0.015 |
| 0.75  | 0.1   | 0.014 | 0.014 | 0.010 | 0.010 | 0.014 | 0.013 | 0.009 | 0.009 |
|       |       | 0.011 | 0.018 | 0.008 | 0.012 | 0.011 | 0.016 | 0.007 | 0.011 |
|       |       | 0.011 | 0.046 | 0.007 | 0.032 | 0.009 | 0.047 | 0.006 | 0.032 |
| 0.5   | 0.5   | 0.013 | 0.013 | 0.009 | 0.009 | 0.013 | 0.012 | 0.009 | 0.008 |
|       |       | 0.010 | 0.018 | 0.007 | 0.013 | 0.010 | 0.017 | 0.006 | 0.012 |
|       |       | 0.008 | 0.045 | 0.005 | 0.032 | 0.006 | 0.045 | 0.004 | 0.032 |
| 0.9   | 0.5   | 0.014 | 0.014 | 0.010 | 0.010 | 0.014 | 0.012 | 0.009 | 0.009 |
|       |       | 0.011 | 0.018 | 0.008 | 0.012 | 0.011 | 0.016 | 0.007 | 0.011 |
|       |       | 0.011 | 0.047 | 0.006 | 0.032 | 0.009 | 0.047 | 0.006 | 0.032 |
Table 4: Standard deviation and asymptotic standard errors of the GEEE estimator with different correlation structures at 3 percentiles with $\rho \in (0.1, 0.5, 0.9)$, and $\varepsilon \sim \mathcal{N}(0,1)$ under a location-scale-shift scenario.

| $\tau$ | $\rho$ | GEEE | SD  | SE  | SD  | SE  | SD  | SE  | SD  | SE  | SD  | SE  |
|------|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.25 | 0.1  | Ind  | 0.026 | 0.021 | 0.019 | 0.017 | 0.022 | 0.020 | 0.017 | 0.015 |
|      |      | AR1  | 0.062 | 0.037 | 0.041 | 0.027 | 0.253 | 0.046 | 0.052 | 0.026 |
|      |      | Un   | 0.048 | 0.054 | 0.033 | 0.040 | 0.043 | 0.054 | 0.032 | 0.045 |
| 0.5  | 0.1  | Ind  | 0.024 | 0.020 | 0.018 | 0.016 | 0.022 | 0.020 | 0.016 | 0.015 |
|      |      | AR1  | 0.060 | 0.038 | 0.038 | 0.024 | 0.241 | 0.055 | 0.048 | 0.025 |
|      |      | Un   | 0.042 | 0.048 | 0.030 | 0.034 | 0.036 | 0.046 | 0.027 | 0.037 |
| 0.9  | 0.1  | Ind  | 0.024 | 0.021 | 0.019 | 0.016 | 0.024 | 0.020 | 0.018 | 0.016 |
|      |      | AR1  | 0.085 | 0.038 | 0.050 | 0.029 | 0.240 | 0.069 | 0.075 | 0.280 |
|      |      | Un   | 0.055 | 0.057 | 0.039 | 0.044 | 0.043 | 0.060 | 0.035 | 0.049 |
| 0.50 | 0.1  | Ind  | 0.027 | 0.023 | 0.019 | 0.017 | 0.024 | 0.020 | 0.017 | 0.016 |
|      |      | AR1  | 0.070 | 0.041 | 0.045 | 0.030 | 0.237 | 0.062 | 0.068 | 0.043 |
|      |      | Un   | 0.053 | 0.055 | 0.036 | 0.039 | 0.059 | 0.065 | 0.038 | 0.048 |
| 0.5  | 0.1  | Ind  | 0.025 | 0.023 | 0.019 | 0.017 | 0.023 | 0.020 | 0.017 | 0.015 |
|      |      | AR1  | 0.064 | 0.038 | 0.040 | 0.028 | 0.239 | 0.063 | 0.063 | 0.043 |
|      |      | Un   | 0.045 | 0.047 | 0.032 | 0.034 | 0.050 | 0.055 | 0.028 | 0.038 |
| 0.9  | 0.1  | Ind  | 0.025 | 0.023 | 0.020 | 0.017 | 0.024 | 0.020 | 0.018 | 0.016 |
|      |      | AR1  | 0.095 | 0.042 | 0.050 | 0.031 | 0.278 | 0.065 | 0.070 | 0.043 |
|      |      | Un   | 0.060 | 0.057 | 0.042 | 0.044 | 0.082 | 0.070 | 0.036 | 0.049 |
| 0.75 | 0.1  | Ind  | 0.028 | 0.024 | 0.020 | 0.018 | 0.028 | 0.024 | 0.020 | 0.017 |
|      |      | AR1  | 0.068 | 0.056 | 0.045 | 0.040 | 0.295 | 0.091 | 0.075 | 0.067 |
|      |      | Un   | 0.062 | 0.053 | 0.038 | 0.038 | 0.059 | 0.062 | 0.042 | 0.049 |
| 0.5  | 0.1  | Ind  | 0.028 | 0.024 | 0.020 | 0.018 | 0.026 | 0.023 | 0.019 | 0.017 |
|      |      | AR1  | 0.057 | 0.052 | 0.039 | 0.037 | 0.296 | 0.086 | 0.065 | 0.063 |
|      |      | Un   | 0.048 | 0.044 | 0.033 | 0.033 | 0.043 | 0.051 | 0.030 | 0.037 |
| 0.9  | 0.1  | Ind  | 0.030 | 0.025 | 0.020 | 0.018 | 0.027 | 0.023 | 0.019 | 0.017 |
|      |      | AR1  | 0.065 | 0.055 | 0.049 | 0.041 | 0.309 | 0.086 | 0.072 | 0.064 |
|      |      | Un   | 0.053 | 0.048 | 0.042 | 0.043 | 0.048 | 0.058 | 0.033 | 0.043 |
Table 5: Total of the frequency 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 |               |               |                 |               |
| \( \mathcal{N} \) | 43     82 550 525 41 61 581 517 | 46 99 759 296 32 49 665 454 |               |               |
| \( T_3 \)     | 98     168 560 374 101 142 587 370 | 96 169 710 225 70 135 702 293 |               |               |
| \( \chi_2^3 \) | 69     138 548 445 60 113 577 450 | 48 130 743 279 49 79 704 368 |               |               |
| Location-scale-shift model |               |               |                 |               |
| \( \mathcal{N} \) | 201    200 245 554 232 185 279 504 | 164 219 325 492 158 194 284 564 |               |               |
| \( T_3 \)     | 99     164 519 418 129 177 466 428 | 125 143 474 458 124 172 442 462 |               |               |
| \( \chi_2^3 \) | 51     115 455 579 41 82 458 619 | 42 112 468 578 53 84 485 578 |               |               |
Table 6: Bias and RMSE 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 $\varepsilon \sim N(0, 1)$.

| $\rho$ | $\tau$ | $n = 50$ |        |        | $n = 100$ |        |        |
|-------|--------|----------|-------|-------|----------|-------|-------|
|       |        | Bias     | RMSE  |       | Bias     | RMSE  |       |
|       |        | Exc lqmm | Exc lqmm | Exc lqmm | Exc lqmm | Exc lqmm | Exc lqmm |
| 0.1   | $\tau_1$ | -0.0005 | -0.0008 | 0.0142 | 0.0154 | -0.0002 | -0.0001 | 0.0104 | 0.0121 |
|       | $\tau_2$ | -0.0006 | -0.0008 | 0.0132 | 0.0147 | 0.0000 | 0.0000 | 0.0099 | 0.0116 |
|       | $\tau_3$ | -0.0006 | -0.0008 | 0.0140 | 0.0161 | 0.0001 | 0.0001 | 0.0102 | 0.0115 |
| 0.5   | $\tau_1$ | 0.0005  | 0.0000 | 0.0130 | 0.0155 | -0.0007 | 0.0000 | 0.0087 | 0.0103 |
|       | $\tau_2$ | 0.0003  | 0.0007 | 0.0122 | 0.0143 | -0.0005 | -0.0005 | 0.0082 | 0.0103 |
|       | $\tau_3$ | 0.0001  | 0.0000 | 0.0128 | 0.0153 | -0.0004 | -0.0002 | 0.0089 | 0.0106 |
| 0.9   | $\tau_1$ | 0.0001  | 0.0002 | 0.0071 | 0.0102 | 0.0003 | -0.0003 | 0.0050 | 0.0067 |
|       | $\tau_2$ | 0.0001  | 0.0000 | 0.0065 | 0.0093 | 0.0004 | 0.0005 | 0.0045 | 0.0069 |
|       | $\tau_3$ | 0.0002  | -0.0001 | 0.0074 | 0.0098 | 0.0005 | 0.0008 | 0.0049 | 0.0072 |

Location-scale-shift model

| $\rho$ | $\tau$ | $n = 50$ |        |        | $n = 100$ |        |        |
|-------|--------|----------|-------|-------|----------|-------|-------|
|       |        | Bias     | RMSE  |       | Bias     | RMSE  |       |
|       |        | Exc lqmm | Exc lqmm | Exc lqmm | Exc lqmm | Exc lqmm | Exc lqmm |
| 0.1   | $\tau_1$ | -0.0008 | -0.0003 | 0.0025 | 0.0243 | -0.0007 | -0.0016 | 0.0020 | 0.0168 |
|       | $\tau_2$ | -0.0001 | 0.0004 | 0.0023 | 0.0189 | 0.0000 | -0.0010 | 0.0017 | 0.0150 |
|       | $\tau_3$ | 0.0006  | -0.0011 | 0.0026 | 0.0223 | 0.0008 | -0.0010 | 0.0020 | 0.0183 |
| 0.5   | $\tau_1$ | -0.0019 | 0.0027 | 0.0033 | 0.0263 | -0.0018 | 0.0023 | 0.0026 | 0.0189 |
|       | $\tau_2$ | -0.0002 | 0.0017 | 0.0026 | 0.0228 | -0.0001 | 0.0011 | 0.0018 | 0.0162 |
|       | $\tau_3$ | 0.0015  | 0.0007 | 0.0031 | 0.0274 | 0.0016 | -0.0006 | 0.0025 | 0.0179 |
| 0.9   | $\tau_1$ | -0.0018 | 0.0155 | 0.0030 | 0.0322 | -0.0017 | 0.0149 | 0.0023 | 0.0249 |
|       | $\tau_2$ | -0.0001 | -0.0001 | 0.0023 | 0.0242 | 0.0000 | 0.0001 | 0.0016 | 0.0174 |
|       | $\tau_3$ | 0.0017  | -0.0150 | 0.0029 | 0.0316 | 0.0017 | -0.0156 | 0.0024 | 0.0256 |
Table 7: Bias and RMSE 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 $\varepsilon \sim \mathcal{N}(0, 1)$.

| $\rho$ | $\tau$ | Location-shift model | Location-scale-shift model |
|-------|--------|----------------------|---------------------------|
|       |        | $n = 50$             | $n = 100$                 |
|       |        | Bias                 | RMSE                      | Bias                 | RMSE                      |
|       |        | Exc lqmm             | Exc lqmm                  | Exc lqmm             | Exc lqmm                  |
| $\tau_1$ | 0.0007 | 0.0010 | 0.0135 0.0160 | -0.0006 -0.0006 | 0.0092 0.0097 |
| $\tau_2$ | 0.0005 | 0.0009 | 0.0129 0.0153 | -0.0008 -0.0009 | 0.0087 0.0095 |
| $\tau_3$ | 0.0001 | 0.0001 | 0.0136 0.0153 | -0.0009 -0.0009 | 0.0093 0.0104 |
| 0.5 $\tau_1$ | -0.0002 | -0.0003 | 0.0114 0.0135 | 0.0002 0.0001 | 0.0084 0.0098 |
| $\tau_2$ | -0.0003 | -0.0006 | 0.0107 0.0127 | 0.0002 0.0001 | 0.0079 0.0093 |
| $\tau_3$ | -0.0004 | 0.0004 | 0.0115 0.0136 | 0.0002 0.0000 | 0.0082 0.0100 |
| 0.9 $\tau_1$ | 0.0005 | 0.0009 | 0.0067 0.0096 | 0.0000 0.0003 | 0.0048 0.0065 |
| $\tau_2$ | 0.0003 | 0.0007 | 0.0061 0.0089 | 0.0000 -0.0006 | 0.0043 0.0058 |
| $\tau_3$ | 0.0000 | 0.0003 | 0.0068 0.0088 | 0.0000 0.0003 | 0.0048 0.0065 |
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 and Bottai (2007), Lu and 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 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

\[
y_{it} = \beta_0 + \beta_1 R_i + \beta_2 T_{it} + \beta_3 R_i T_{it} + \varepsilon_{it},
\]

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 30 min. The corresponding GEEE equation, for a fixed \(\tau\)

\[
\mu_\tau(y_{it}) = \beta_{0\tau} + \beta_1 R_i + \beta_2 T_{it} + \beta_3 R_i T_{it},
\]

with fourth working correlation (Ind, Exc, AR1, Un), was estimated for three asymmetric points, \((0.25, 0.5, 0.75)\). Table 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 \(\hat{\beta}_1\) is not significant in any of the models and the estimated parameter \(\hat{\beta}_0\) 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 \(\hat{\beta}_2\) and \(\hat{\beta}_3\) are significant at 5% level for the GEEE methods, except for \(\hat{\beta}_3\) of the GEEE with an Exc working correlation. This means that time and its interaction with treatment affect the amount 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. We see that medication helps women relieve 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 \((0.25, 0.5, 0.75)\).

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 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.
Table 8: Parameters estimates (Est) with their standard errors (SE) and 95% confidence interval (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** |          |        |        |         |        |        |         |         |        |
| Ind    |           |        |        |         |        |        |         |         |        |
| $\beta_0$ | 2.63     | 15.66  | 35.76  | 4.83    | 6.62   | 8.06   | (-6.83, 12.09) | (2.68, 28.64) | (19.97, 51.55) |
| $\beta_1$ | 4.34     | -2.23  | -12.92 | 5.37    | 7.69   | 9.89   | (-6.19, 14.86) | (-17.31, 12.85) | (-32.30, 6.46) |
| $\beta_2$ | 10.70    | 11.33  | 9.84   | 1.97    | 1.62   | 1.48   | (6.83, 14.57)  | (8.16, 14.49)  | (6.94, 12.74)  |
| $\beta_3$ | -9.65    | -9.58  | -7.32  | 2.12    | 2.03   | 2.22   | (-13.80, -5.49) | (-13.56, -5.59) | (-11.67, -2.97) |
| **Exc** |           |        |        |         |        |        |         |         |        |
| $\beta_0$ | 1.69     | 17.14  | 38.17  | 11.47   | 12.61  | 12.06  | (-20.78, 24.17) | (-7.59, 41.86) | (14.53, 61.81) |
| $\beta_1$ | 1.68     | 14.31  | 12.19  | 12.50   | 14.56  | 14.22  | (-22.83, 26.18) | (-32.85, 24.22) | (-40.07, 15.69) |
| $\beta_2$ | 10.15    | 11.25  | 10.05  | 4.65    | 4.31   | 3.61   | (1.03, 19.26)   | (2.80, 19.70)   | (2.97, 17.14)   |
| $\beta_3$ | -8.71    | -9.15  | -7.03  | 4.94    | 4.96   | 4.34   | (-18.39, 0.96)  | (-18.87, 0.56)  | (-15.53, 1.47) |
| **AR1** |           |        |        |         |        |        |         |         |        |
| $\beta_0$ | 3.53     | 15.80  | 33.07  | 6.37    | 8.92   | 10.98  | (-8.95, 16.02)  | (-1.68, 33.28)  | (11.55, 54.59)  |
| $\beta_1$ | 4.29     | -1.69  | -13.22 | 7.44    | 10.83  | 14.85  | (-10.28, 18.87) | (-22.92, 19.54) | (-42.33, 15.89) |
| $\beta_2$ | 8.06     | 9.08   | 8.76   | 1.37    | 1.57   | 1.59   | (5.37, 10.74)   | (6.01, 12.16)   | (5.65, 11.87)   |
| $\beta_3$ | -7.62    | -8.14  | -6.52  | 1.64    | 2.21   | 2.91   | (-10.84, -4.40) | (-12.46, -3.81) | (-12.23, -0.82) |
| **Un**  |           |        |        |         |        |        |         |         |        |
| $\beta_0$ | 2.24     | 16.95  | 38.67  | 6.36    | 7.95   | 9.01   | (-10.23, 14.70) | (1.36, 32.54)   | (21.00, 56.34)  |
| $\beta_1$ | 2.50     | -4.15  | -12.36 | 6.96    | 9.14   | 10.56  | (-11.15, 16.14) | (-22.07, 13.76) | (-33.06, 8.33)  |
| $\beta_2$ | 10.39    | 11.30  | 10.05  | 2.81    | 2.52   | 2.24   | (4.88, 15.89)   | (6.36, 16.25)   | (5.65, 14.45)   |
| $\beta_3$ | -9.13    | -9.21  | -6.95  | 2.97    | 2.97   | 2.94   | (-14.94, -3.32) | (-15.03, -3.39) | (-12.70, -1.19) |
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 and Fan 2015, Leng and 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: Representation of the estimated labor pain obtained from the GEE method with the different working correlations (Ind, Exc, AR1, Un) at three percentiles, (0.25, 0.5, 0.75).
Sketches of the proofs are provided below. A detailed version is available in the supplementary material II.

**Proof of Theorem 1**

The proof of Theorem 1 is based on the following result from Hjort (2011).

**Corollary 1.** (Basic Corollary of Hjort (2011)) Consider $A_n(s)$ is convex and can be represented as $\frac{1}{2} s^T V s + U_n^T s + C_n + r_n(s)$, where $V$ is symmetric and positive definite, $U_n$ is stochastically bounded, $C_n$ is arbitrary, and $r_n(s)$ goes to zero in probability for each $s$. Then $\alpha_n$ the argmin of $A_n$, is only $o_p(1)$ away from $\beta_n = -V^{-1}U_n$, the argmin of $\frac{1}{2} s^T V s + U_n^T s + C_n$. If also $U_n \xrightarrow{d} U$, then $\alpha_n \xrightarrow{d} -V^{-1}U_n$.

With this result, the proof of Theorem 1 is obtained by approximating the objective function $R_{Nq}$ using a convex quadratic function with a unique minimum. This approximation is possible through the following Lemma:

**Lemma 1.** Let $r(t) = \rho_{\tau}(u-t) - \rho_{\tau}(t) + 2t\psi_{\tau}(u)\tau$ then

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

**Proof of Lemma 1**

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

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

According to the sign of $t$, 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 $t > 0$, and

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

if $t < 0$. Hence we have $r(t) = O(t^2)$.

The objective function $R_{Nq}$ is defined as

$$R_{Nq}(\delta) = \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{m_i} w_k \left\{ \rho_{\tau}(\varepsilon_{it\tau_k} - x_{it\tau_k}^T \delta_{\tau_k}/\sqrt{N}) - \rho_{\tau}(\varepsilon_{it\tau_k}) \right\}$$  \hspace{1cm} (22)

where $\delta = (\delta_{\tau_1}, \ldots, \delta_{\tau_q})^T$ is a $pq \times 1$ vector, $\delta_{\tau_k}^T$ a $p \times 1$ vector of parameter and $\varepsilon_{it\tau_k} = y_{it} - \mu_{it\tau_k}$ and $\mu_{it\tau_k} = x_{it\tau_k}^T \beta_{\tau_k}$.

$R_{Nq}$ is a convex function of $\delta$ and is minimized by $\hat{\delta}$.
\[
\hat{\delta} = \begin{pmatrix}
\delta_{\tau_1} \\
\vdots \\
\delta_{\tau_q}
\end{pmatrix} = \begin{pmatrix}
\sqrt{N}(\beta_{\tau_1} - \beta_{\tau_1}) \\
\vdots \\
\sqrt{N}(\beta_{\tau_q} - \beta_{\tau_q})
\end{pmatrix}
\]

Now using Lemma 1 we are able to split the objective function in two functions \( R_{Nq}^{(1)} \) and \( R_{Nq}^{(2)} \)

\[
R_{Nq}(\delta) \simeq -2 \frac{1}{\sqrt{N}} \sum_{k=1}^{q} w_k \delta_{\tau_k}^T X^T \Psi_{\tau_k}(\varepsilon_{\tau_k}) \varepsilon_{\tau_k} + \frac{1}{N} \sum_{k=1}^{q} w_k \delta_{\tau_k}^T X^T \mathbb{E}[\Psi_{\tau_k}(\varepsilon_{\tau_k})]X \delta_{\tau_k}
\]

\[
= R_{Nq}^{(1)}(\delta) + R_{Nq}^{(2)}(\delta).
\]

Conditions A2 and A3 imply a Lindeberg condition and we have

\[
R_{Nq}^{(1)}(\delta) = \frac{-2}{\sqrt{N}} \delta^T (W \otimes X)^T \Psi(\varepsilon) \varepsilon \xrightarrow{d} -2\delta^T B
\]

where \( B \) is a zero mean Gaussian vector with covariance matrix \( D_{I0}(\tau) \). For the second term to the right of \( R_{Nq} \) we have, by condition A2,

\[
R_{Nq}^{(2)}(\delta) = N^{-1} \delta^T \sum_{i=1}^{n} (W \otimes X_i)^T \mathbb{E} [\Psi(\varepsilon) \varepsilon] (I_q \otimes X_i) \delta
\]

\[
\xrightarrow{d} \delta^T D_{I1}(\tau) \delta.
\]

Thus the limiting form of the objective function is

\[
R_{0q}(\delta) = -2\delta^T B + \delta^T D_{I1}(\tau) \delta
\]

where \( B \) is a zero mean Gaussian vector with covariance matrix \( D_{I0}(\tau) \). Application of Corollary 1 gives the result of Theorem 1.

**Proof of Theorem 2**

For the proof of Theorem 2 we must show the convergence of \( \hat{D}_{I1}(\tau) \) and \( \hat{D}_{I0}(\tau) \).

The matrix \( \hat{D}_{I1}(\tau) \) is a diagonal matrix of general term \( N^{-1} w_k \sum_{i=1}^{n} X_i^T \Psi_{\tau_k}(\hat{\varepsilon}_{\tau_k})X_i, 1 < k < q \). It suffices to show the convergence of \( N^{-1} \sum_{i=1}^{n} X_i^T \Psi_{\tau_k}(\hat{\varepsilon}_{\tau_k})X_i \) to obtain that of \( \hat{D}_{I1}(\tau) \). Using the result \( \text{plim} \hat{\beta}_{\tau_k} = \beta_{\tau_k} \), the convergence of \( \hat{D}_{I1}(\tau) \xrightarrow{d} D_{I1}(\tau) \) follows by application of the Markov’s inequality and the Markov law of large numbers (LLN).

The matrix \( \hat{D}_{I0}(\tau) \) is a block matrix of dimension \( pq \times pq \) and of general term \( w_k w_j N^{-1} \sum_{i=1}^{n} X_i^T \hat{\Sigma}_{i\tau_k\tau_j}X_i, 1 < k, j < q \). To obtain the convergence of \( \hat{D}_{I0}(\tau) \), it suffices to show the convergence of its general term. The convergence of the general term follows from repeated use of the Cauchy–Schwarz inequality, Minkowski’s inequality and the Markov LLN.
Proof of Theorem 3. By application of the Markov LLN, we have $N^{-1}\{S(\beta) - \mathbb{E}[S(\beta)]\} \xrightarrow{p} o_p(1)$. Since $S(\beta)$ is a convex function of $\beta$, the convergence is uniform by the convexity Lemma of Pollard (1991). $\mathbb{E}[S(\beta)]$ has a unique minimum and by Lemma A of Newey (1987), we have $\hat{\beta}$, the solution of $S(\beta) = 0$ converges in probability to $\beta$, unique solution of $\mathbb{E}[S(\beta)]$, that is, $\hat{\beta} \xrightarrow{p} \beta$.

Using condition B2 and application of Liapounov CLT, we can show that $N^{-1/2}S(\beta)$ is a zero mean Gaussian vector with covariance matrix $D_0(\tau)$.

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

$$N^{-1}S(\beta) = N^{-1}S(\beta) + N^{-1}D_1(\beta)(\beta - \beta) + N^{-1}\left[-\sum_{i=1}^{n}(W \otimes X_i)^T V_i^{-1}\Psi(\epsilon)(I_q \otimes X_i) - D_1(\beta)\right](\beta - \beta).$$

The last term of the equation is $o_p(1)$ by application of the LLN and the Slutsky’s Theorem. Because $S(\hat{\beta}) = 0$ and $\hat{\beta}$ is in the neighborhood of $\beta$, we have

$$\sqrt{N}(\hat{\beta} - \beta) = -D_1(\tau)\frac{1}{\sqrt{N}}S(\beta) + o_p(1).$$

Therefore $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, D_1^{-1}(\tau)D_0(\tau)D_1^{-1}(\tau))$. \hfill \Box

Proof of Theorem 4.

As for Theorem 2, the proof of Theorem 4 consists in showing the convergence of $\hat{D}_1(\tau)$ and $\hat{D}_0(\tau)$. This is done following the same steps as Theorem 2. \hfill \Box
References

Aigner, D.J., T. Amemiya, and D.J. Poirier. 1976. “ON the Estimation of Production Frontiers: MAXIMUM Likelihood Estimation of the Parameters of a Discontinuous Density Function.” *International Economic Review* 17 (2): 377.

Azzalini, A. 1985. “A Class of Distributions Which Includes the Normal Ones.” *Scandinavian Journal of Statistics* 12 (2). [Board of the Foundation of the Scandinavian Journal of Statistics, Wiley]: 171–78. http://www.jstor.org/stable/4615982

Chen, Li, Lee-Jen Wei, and Michael I. Parzen. 2004. “Quantile Regression for Correlated Observations.” In *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, edited by D. Y. Lin and P. J. Heagerty, 51–69. New York, NY: Springer New York. doi:10.1007/978-1-4419-9676-1_4

Davis, Charles S. 1991. “Semi-Parametric and Non-Parametric Methods for the Analysis of Repeated Measurements with Applications to Clinical Trials.” *Statistics in Medicine* 10 (12). Wiley Subscription Services, Inc., A Wiley Company: 1959–80. doi:10.1002/sim.4780101210

Diggle, P., P. Heagerty, K.Y. Liang, and S. Zeger. 2013. *Analysis of Longitudinal Data*. Oxford Statistical Science Series. OUP Oxford. https://books.google.ca/books?id=zAiK-gWUqDUC

Efron, B. 1991. “REGRESSION Percentiles Using Asymmetric Squared Error Loss.” *Statistica Sinica* 1 (1). Institute of Statistical Science, Academia Sinica: 93–125.

Eilers, Paul HC. 2013. “Discussion: The Beauty of Expectiles.” *Statistical Modelling* 13 (4): 317–22.

Farooq, M., and I. Steinwart. 2017. “An Svm-Like Approach for Expectile Regression.” *Computational Statistics and Data Analysis* 109. Elsevier B.V.: 159–81. doi:10.1016/j.csda.2016.11.010

Fitzmaurice, G., M. Davidian, G. Verbeke, and G. Molenberghs. 2008. *Longitudinal Data Analysis*. Chapman & Hall/Crc Handbooks of Modern Statistical Methods. CRC Press. https://books.google.ca/books?id=zVBjCvQCoGQC

Fu, Liya, and You-Gan Wang. 2012. “Quantile Regression for Longitudinal Data with a Working Correlation Model.” *Computational Statistics & Data Analysis* 56 (8): 2526–38. doi:https://doi.org/10.1016/j.csda.2012.02.005

Furlotte, Nicholas A., Eleazar Eskin, and Susana Eyheramendy. 2012. “Genome-Wide Association Mapping with Longitudinal Data.” *Genetic Epidemiology* 36 (5): 463–71.

Geraci, M., and M. Bottai. 2007. “Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution.” *Biostatistics* 8 (1): 140–54.

Geraci, Marco, and Matteo Bottai. 2014. “Linear Quantile Mixed Models.” *Statistics and Computing* 24 (3): 461–79. doi:10.1007/s11222-013-9381-9

Hsiao, Cheng. 2007. “Panel Data Analysis—advantages and Challenges.” *TEST* 16 (1): 1–22.

Jang, Mi. 2011. “Working Correlation Selection in Generalized Estimating Equations.” ProQuest Dissertations Publishing. http://search.proquest.com/docview/920555336/

Jiang, C., M. Jiang, Q. Xu, and X. Huang. 2017. “Expectile Regression Neural Network Model with Applications.” *Neurocomputing* 247. Elsevier B.V.: 73–86. doi:10.1016/j.neucom.2017.03.040

Jung, Sin-Ho. 1996. “Quasi-Likelihood for Median Regression Models.” *Journal of the American Statistical Association* 91 (433). Taylor & Francis: 251–57.

Kim, M., and S. Lee. 2016. “Nonlinear Expectile Regression with Application to Value-at-Risk and Expected Shortfall Estimation.” *Computational Statistics and Data Analysis* 94. Elsevier: 1–19. doi:10.1016/j.csda.2015.07.011

Kneib, Thomas. 2013a. “Beyond Mean Regression.” *Statistical Modelling* 13 (4): 275–303.
and Goliath?” *Statistical Modelling* 15 (5): 433–56. doi:10.1177/1471082X14561155

White, Halbert. 1980. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica* 48 (4): 817–38. https://EconPapers.repec.org/RePEc:ecm:emetrp:v:48:y:1980:i:4:p:817-38

Xing, J.-J., and X.-Y. Qian. 2017. “Bayesian Expectile Regression with Asymmetric Normal Distribution.” *Communications in Statistics - Theory and Methods* 46 (9). Taylor; Francis Inc.: 4545–55. doi:10.1080/03610926.2015.1088030

Xu, Q., X. Liu, C. Jiang, and K. Yu. 2016. “Nonparametric Conditional Autoregressive Expectile Model via Neural Network with Applications to Estimating Financial Risk.” *Applied Stochastic Models in Business and Industry* 32 (6). John Wiley; Sons Ltd: 882–908. doi:10.1002/asmb.2212

Yang, Y., and H. Zou. 2015. “Nonparametric Multiple Expectile Regression via Er-Boost.” *Journal of Statistical Computation and Simulation* 85 (7). Taylor; Francis Ltd.: 1442–58. doi:10.1080/00949655.2013.876024