Parametric Modeling of Quantile Regression Coefficient Functions With Longitudinal Data

Paolo Frumento\textsuperscript{a}, Matteo Bottai\textsuperscript{b}, and Iván Fernández-Val\textsuperscript{c}

\textsuperscript{a}Department of Political Sciences, University of Pisa, Pisa, Italy; \textsuperscript{b}Unit of Biostatistics, Karolinska Institutet, Institute of Environmental Medicine, Stockholm, Sweden; \textsuperscript{c}Department of Economics, Boston University, Boston, MA

**ABSTRACT**

In ordinary quantile regression, quantities of different order are estimated one at a time. An alternative approach, which is referred to as \textit{quantile regression coefficients modeling} (qrcm), is to model quantile regression coefficients as parametric functions of the order of the quantile. In this article, we describe how the qrcm paradigm can be applied to longitudinal data. We introduce a two-level quantile function, in which two different quantile regression models are used to describe the (conditional) distribution of the within-subject response and that of the individual effects. We propose a novel type of penalized fixed-effects estimator, and discuss its advantages over standard methods based on $\ell_1$ and $\ell_2$ penalization. We provide model identifiability conditions, derive asymptotic properties, describe goodness-of-fit measures and model selection criteria, present simulation results, and discuss an application. The proposed method has been implemented in the R package qrcm.

**1. Introduction**

Quantile regression (e.g., Koenker and Bassett 1978; Koenker 2005) has become a standard method in many fields, including medicine, epidemiology, economics, and social sciences. Different solutions have been proposed to extend quantile regression to longitudinal data, in which the same individuals or clusters are observed repeatedly.

In conditional models, that include fixed- and random-effects models, the dependence between observations is accounted for by introducing individual-specific parameters, or “individual effects.” In fixed-effects models, the individual effects are treated as parameters, avoiding distributional assumptions and allowing for a simple computation. A penalized fixed-effects estimator for longitudinal quantile regression has been proposed by Koenker (2004), and similar approaches have been used in Lamarche (2010), Canay (2011), and Kato, Galvao, and Montes-Rojas (2012).\textsuperscript{1} In random-effects models, the individual effects are described by a parametric distribution. Different methods have been proposed to combine the parametric likelihood of the random effects with the estimating equation of ordinary quantile regression. Geraci and Bottai (2007, 2014) used the log-likelihood of an asymmetric Laplace distribution, and Kim and Yang (2011) described an empirical likelihood method. Abrevaya and Dahl (2008) adapted the correlated random effect approach of Chamberlain (1984) to quantile regression, and Arellano and Bonhomme (2016) marginalized the loss function of quantile regression with respect to the posterior distribution of the individual effects. Farcomeni (2012), Marino, Tzavidis, and Alfó (2018), and Alfó, Salvati, and Ranalli (2017) used finite mixtures to approximate the probability density function of the individual effects through a discrete distribution.

Marginal models have also been described in the literature. Leng and Zhang (2014) defined a set of unbiased estimating equations carrying information on the correlation structure. A similar approach was used by Zhao, Lian, and Liang (2017) to implement longitudinal single-index quantile regression.

In this article, we adopt the conditional paradigm and introduce a two-level quantile function, in which both the distribution of the within-subject response (level 1) and that of the individual effects (level 2) are described by quantile regression models. With this approach, the distribution of the individual effects is not subject to strong parametric assumptions and is allowed to depend on level-2 covariates. Following Frumento and Bottai (2016, 2017) and Yang, Chen, and Chang (2017), we describe the level-1 and level-2 quantile regression coefficients by (flexible) parametric functions of the order of the quantile. Compared with standard quantile regression, in which quantiles are estimated one at a time, this modeling approach presents numerous advantages, that include a simpler computation and inference (owing to a smooth objective function), increased statistical efficiency, and easy interpretability of the results.

To fit the model, we introduce a new form of penalized fixed-effects estimator in which the penalty term carries information on level-2 parameters. This method presents important advantages over standard $\ell_1$ and $\ell_2$ penalization. In particular, it avoids the problem of selecting a tuning constant, and allows to

---

\textsuperscript{1}Chernozhukov, Fernández-Val, and Weidner (2018) considered an alternative to quantile regression for estimation of quantile effects in longitudinal data based on distribution regression.
2. The Model

2.1. A Two-Level Quantile Function

We consider a cluster data structure, in which $N$ individuals or clusters are observed repeatedly. We denote by $i = 1, \ldots, N$ the index of the subject, and by $t = 1, \ldots, T$ the within-subject index, such that the total sample size is $NT$. Designs in which $T$ varies across clusters are also possible, at the cost of a slightly more complicated notation.

We denote by $Y_{it}$ a response variable of interest, and assume that

$$
y_{it} = x_{it}^T \beta(U_{it}) + z_{it}^T \gamma(V_{it}),$$

where $x_{it}$ is a $d_x$-dimensional vector of level-1 covariates, with associated parameter $\beta(\cdot)$; and $z_{it}$ is a $d_z$-dimensional vector of level-2 covariates, with associated parameter $\gamma(\cdot)$.

We assume that (i) $x_{it}^T \beta(\cdot)$ and $z_{it}^T \gamma(\cdot)$ are a.s. nondecreasing functions of their arguments, and (ii) $U_{it}$ and $V_{it}$ are $U(0, 1)$ variables, independent of each other and of the covariates. Based on model (1), $\alpha_i = z_{it}^T \gamma(V_{it})$ is an individual effect with conditional quantile function $z_{it}^T \gamma(\cdot)$, while $x_{it}^T \beta(\cdot)$ is the conditional quantile function of $Y_{it} - \alpha_i$.

The level-1 quantile regression model, $x_{it}^T \beta(\cdot)$, has the standard interpretation (e.g., Koenker 2004): it characterizes the “within” part of the distribution, purged of the individual effects. The level-2 regression model, $z_{it}^T \gamma(\cdot)$, describes the distribution of the between-subject differences with respect to a reference value which typically corresponds to a “mean” or “median” individual.

Consider, for example, a clinical study in which patients are repeatedly measured their body mass index (BMI) during their lifetime. The level-1 part of the model describes the conditional quantiles of BMI in a “typical” patient, that is, someone with an individual effect equal to 0. Level-1 predictors include time-varying characteristics, such as the age of the patient at each observation, as well as constant traits, such as the gender of the patient. The level-2 model accounts for the between-patient heterogeneity, and describes the conditional quantiles of the individual effects. Level-2 covariates can only include time-invariant traits, such as the gender, and summary statistics of level-1 covariates, for example, the age at the first examination. Note that the dimension of the level-1 covariates, $x_{it}$, is $NT$, while that of the level-2 covariates, $z_{it}$, is $N$.

Unlike the “standard” approaches, that do not consider the effect of level-2 covariates, our modeling framework allows to investigate the determinants of the between-subject variability. For example, in the linear random-intercept model, the level-2 response is described by a $N(0, \phi^2)$ distribution, in which $\phi^2 = \text{var}(\alpha_i)$ is interpreted as the “between” variance and is assumed to be unaffected by predictors. This model may fail to capture important features of the data, such as the fact that the variance of the individual effects is different in males and females. Model (1), instead, allows including gender as level-2 predictor.

Using a quantile regression approach permits avoiding strong parametric assumptions such as normality and homoscedasticity that are often used in likelihood-based modeling. In the existing literature on longitudinal quantile regression, however, a quantile regression model is usually applied to the level-1 response, but not to the individual effects, that are treated as nuisance parameters. In our paradigm, instead, the two parts of the distribution are considered “equally important,” in the sense that the same modeling structure is used to describe the quantiles of the between-subject response, and those of the between-subject differences. As shown later in the article, working with model (1) permits using the same techniques to estimate both the level-1 and the level-2 parameters, and avoids combining level-1 quantile regression methods with likelihood-based level-2 estimators as for example in Kim and Yang (2011). This leads to rather simple procedures for estimation and inference, in which a fundamental role is played by the two independent uniform random variables ($U_{it}$, $V_{it}$) that generate the data.

2.2. Parametric Coefficient Functions

Through the article, we assume that the quantile regression coefficient functions, $\beta(\cdot)$ and $\gamma(\cdot)$, can be modeled parametrically:

$$
\beta(u) = \beta(u|\theta), \quad \gamma(v) = \gamma(v|\phi),
$$

where $\theta$ and $\phi$ are unknown model parameters. This modeling approach was used by Frumento and Bottai (2016, 2017), and is exemplified in Figure 1. The broken line in the figure represents standard regression coefficients at quantiles $u = (0.01, 0.02, \ldots, 0.99)$. The estimated coefficients show a nonsmooth, volatile trend and, although consistently positive, are almost never significant. A parametric model can be used to characterize the coefficient function with few parameters and describe it by a simple, closed-form mathematical expression. In Figure 1, we propose a linear fit, $\beta(u|\theta) = \theta_0 + \theta_1 u$, that is represented by a dashed line. This simple model reveals the underlying trend and permits achieving statistical significance.

Compared with standard quantile regression, which works in a quantile-by-quantile fashion, modeling quantile functions parametrically simplifies estimation and inference and yields important advantages in terms of parsimony, efficiency, and ease of interpretation. Moreover, it allows for model identification in the presence of latent structures or missing information, making
it simple to apply quantile regression to censored and truncated data (Frumento and Bottai 2017).

On the other hand, this approach requires formulating a parametric model for the coefficient functions, $\beta(u|\theta)$ and $\gamma(v|\phi)$. This task is not straightforward and the existing literature on the subject is lacking. In Section 3, we describe in details model building, provide guidelines, and suggest a variety of possible parameterizations.

3. Two-Level Modeling of Quantile Regression Coefficient Functions

We assume model (1) to hold, and parameterize the quantile regression coefficient functions as follows:

$$
\beta(u|\theta) = \theta b(u), \quad \gamma(v|\phi) = \phi c(v),
$$

(3)

where $b(u) = [b_1(u), \ldots, b_d(u)]^T$ and $c(v) = [c_1(v), \ldots, c_d(v)]^T$ are $d_b$- and $d_c$-dimensional sets of known functions. With this notation, $\theta$ is a $d_x \times d_b$ matrix, and $\phi$ is a $d_x \times d_c$ matrix. The data-generating process can be written as

$$
Y_{it} = x_{it}^T \theta b(U_{it}) + z_{it}^T \phi c(V_i).
$$

(4)

Although other parameterizations are possible (e.g., $\beta(u|\theta)$ and $\gamma(v|\phi)$ may be allowed to be nonlinear functions of $\theta$ and $\phi$), model (3) is very flexible and computationally convenient. We illustrate the potentials of this modeling approach with a number of examples, and provide general guidelines for model building.

3.1. A Simple Model

Consider the following model with a single level-1 covariate $x$, and no level-2 predictors:

$$
Y_{it} = \beta_0(U_{it}) + \beta_1(U_{it}) x_{it} + \gamma_0(V_i).
$$

Denote by $\zeta(\cdot)$ the quantile function of a standard normal distribution, and assume that

$$
\beta_0(u|\theta) = \theta_{00} + \theta_{01} \zeta(u),
$$

$$
\beta_1(u|\theta) = \theta_{10},
$$

$$
\gamma(v|\phi) = \phi \zeta(v).
$$

This is just a reformulation of the standard linear random-intercept model, in which $Y_{it} = \theta_{00} + \theta_{01} x_{it} + \alpha_i + \epsilon_{it}$ with $\alpha_i \sim N(0, \phi^2)$ and $\epsilon_{it} \sim N(0, \phi^2)$. In this model, $\theta_{00}$ corresponds to the intercept of the “fixed” part, while $\theta_{01}$ and $\phi^2$ are interpreted as the “within” and “between” variance components. In the equivalent quantile regression model, $\theta_{00}$ is the “intercept” of $\beta_0(u|\theta)$ and corresponds to $\beta_0(0.5|\theta)$, while $\theta_{01}$ and $\phi$ are “slopes” associated with $\zeta(\cdot)$ in the level-1 and level-2 part of the quantile function, respectively. The regression coefficient of $x$, $\beta_1(u|\theta)$, is assumed to be constant across quantiles, forcing homoscedasticity.

3.2. A More Flexible Model

The standard linear random-intercept model is rather restrictive and, within the described framework, can be easily generalized by choosing a different specification of $\beta(\cdot|\theta)$ and $\gamma(\cdot|\phi)$. For example, one may define

$$
\beta_0(u|\theta) = \theta_{00} + \theta_{01} u + \theta_{02} u^2 + \theta_{03} u^3 + \theta_{04} \zeta(u),
$$

$$
\beta_1(u|\theta) = \theta_{10} + \theta_{11} u,
$$

$$
\gamma(v|\phi) = \phi_1 \log(2v) + \phi_2 \log(2(1 - v)).
$$

Although other parameterizations are possible (e.g., $\beta(u|\theta)$ and $\gamma(v|\phi)$ may be allowed to be nonlinear functions of $\theta$ and $\phi$), model (3) is very flexible and computationally convenient. We illustrate the potentials of this modeling approach with a number of examples, and provide general guidelines for model building.

As shown in this example, $\beta(\cdot)$ and $\gamma(\cdot)$ can be constructed as linear combinations of relatively simple functions, $b(\cdot)$ and $c(\cdot)$, such that $\beta(u|\theta) = \theta b(u)$, and $\gamma(v|\phi) = \phi c(v)$. In this framework, the model is entirely determined by the choice of $b(\cdot)$ and $c(\cdot)$. Useful guidelines for model building are provided in the rest of this section. Various modeling approaches are illustrated in Sections 7 and 8 of this article, while a general discussion on quantile modeling can be found in the book by Gilchrist (2000). Finally, the documentation of the qcr package (in particular the functions `iqzr` and `iqzrl`) includes an extensive tutorial for the practitioners.
### 3.3. Modeling β₀(u|θ)

Assuming that the support of x includes the zero (which can be obtained by centering the covariates), β₀(·|θ) must be a monotonically increasing function. Prior belief or knowledge can be used to identify a meaningful parametric model. For instance, one may use the quantile function of a known distribution. Possible parameterizations of β₀(u|θ) include: θ₀₀ + θ₀₁ζ(u), the normal distribution, N(θ₀₀, θ₀₁²); −θ₀₁ log(1 − u), the exponential distribution, Exp(θ₀₁); θ₀₀ + θ₀₁ log(u/(1 − u)), the logistic distribution, Logis(θ₀₀, θ₀₁); θ₀₀ + θ₀₁ log(u) + θ₀₂ log(1 − u), the asymmetric logistic, aLogis(θ₀₀, θ₀₁, θ₀₂); θ₀₀ + θ₀₁ u, the uniform distribution, U(θ₀₀, θ₀₀). Note that, in this framework, the parameters of well-known distributions may have an unusual interpretation. For example, the value of θ₀₁ in a U(θ₀₀, θ₀₀ + θ₀₁) distribution corresponds to its range, but can also be seen as the slope of a linear quantile function, θ₀₀ + θ₀₁ u.

### 3.3.2. Modeling β₁(u|θ), β₂(u|θ), ...

There are no general constraints to the parametric form of the regression coefficients associated with the covariates. However, the coefficient functions are usually bounded and exhibit a rather simple behavior. Sometimes, it is possible to assume that covariates only affect the location of the level-1 response, and force homoscedasticity by choosing a constant-slope model in which β₁(u|θ) = θ₀₁₁ = 1, 2, …, in a more general scenario, a useful approximation is often given by a linear-slope model, β₁(u|θ) = θ₀₁ + θ₁₁ u, or a quadratic-slope model, β₁(u|θ) = θ₀₁ + θ₁₁ u + θ₁₂ u², which does not impose monotone effect with respect to u.

### 3.4. Model Building: Level 2

A similar model strategy can be applied to the level-2 quantile function. There are, however, some important differences.

#### 3.4.1. Modeling γ₀(v|φ)

The distribution of the individual effects is typically assumed to have zero mean or median, and, for identifiability, γ₀(v|φ) does not usually include a constant term. Meaningful definitions of γ₀(v|φ) include: φ₀₁ζ(v), the normal distribution, N(0, φ₀₁²); −φ₀₁ log(1 − v), the exponential distribution, Exp(φ₀₁); φ₀₁ log(v/(1 − v)), the logistic distribution, Logis(0, φ₀₁); φ₀₁ log(2v) + φ₀₂ log(2(1 − v)), a zero-median asymmetric logistic; φ₀₁ log(v) + 1 + φ₀₂ log(1 − v) + 1, a zero-mean asymmetric logistic; φ₀₁(2v − 0.5), a centered uniform distribution, U(−φ₀₁, φ₀₁). In most cases, the coefficients can be interpreted as scale parameters, while the centrality parameter is fixed and equal to zero. In the exponential case, the value 0 is the minimum of the support of the individual effects, and not a measure of central tendency, while both the mean and the standard deviation of the individual effects correspond to φ₀₁.

#### 3.4.2. Modeling γ₁(v|φ), γ₂(v|φ), ...

Importantly, the described framework permits investigating how the conditional quantile function of the individual effects depends on level-2 covariates zᵢ, which typically include cluster-

### 3.5. Additional Remarks

The problem of formulating a parametric quantile function is equivalent, at least in principle, to that of choosing a parametric form for a probability density function, a hazard function, or a survival function. For example, as shown in Section 3.1, standard parametric assumptions such as normality and homoscedasticity can be directly translated into a quantile function with a simple closed-form expression. However, as suggested in Section 3.2, the models that can be used to describe a quantile function are often very different from most of the "conventional" parametric distributions, and frequently much more flexible.

An exploratory semiparametric fit can be obtained by letting b(·) and c(·) be the basis of a linear or polynomial spline. A flexible model can be used as a guide to find more parsimonious and efficient parameterizations. Note that standard quantile regression, in which quantiles are estimated one at a time, can be thought of as a model in which b(·) and c(·) are allowed to be arbitrarily flexible and the parameters θ and φ are virtually infinite-dimensional.

In absence of prior knowledge, one may define b(·) and c(·) using polynomials [e.g., u, u², u³, …], roots [e.g., u¹/², (1 − u)¹/², u¹/³, (1 − u)¹/³, …], trigonometric functions [e.g.,
cos(2\pi u), \sin(2\pi u)], splines, and combinations of the above. A possible strategy is to consider a "simple" quantile function (e.g., that of a normal or an exponential distribution, depending on the nature of the outcome) and allow for a departure from it, as suggested in Section 3.2.

Importantly, the model specification should reflect assumptions on the shape, support, and boundedness (or unboundedness) of the level-1 and level-2 responses. For example, if the individual effects are believed to be symmetric, \( \gamma_0(\psi) \) could be formed by the quantile function of a normal or logistic distribution. If the level-1 distribution has a long right tail, \( \beta_0(u|\theta) \) may have a positive asymptote in \( u = 1 \), for example, \( \beta_0(u|\theta) = \theta_0 + \theta_1 \log(1 - u) + \cdots \). On the other hand, if the outcome is strictly positive, building blocks such as \( \log(u) \) or \( \zeta(u) \), that present a negative asymptote in \( u = 0 \), may not be appropriate.

Apart from the above important considerations, the choice of \( b(\cdot) \) and \( c(\cdot) \) is not as crucial as it appears. For example, the coefficient function defined by \( \beta(u) = (u - 0.3)^3 \) is almost identical to \( \beta(u) = -1.87 + 6.20 u + 1.84 \cos(u) - 5.92 \sin(u) \), the correlation between the two being about 0.99999. The fact that very different model specifications can be indistinguishable in terms of model fit is unsurprising (e.g., it is almost impossible to distinguish a Normal distribution, a Student’s t distribution with large degrees of freedom, and a Gamma distribution with large shape parameter), and suggests that meaningful criteria for model selection should include parsimony and interpretability.

Often, a rather restrictive model may provide a reasonable approximation of the true data distribution, and can be preferred to a more complex model because of its simplicity. Also, parsimonious models are very rewarding in terms of precision, although they may introduce some bias. This explains why strong parametric assumptions, such as homoscedasticity and proportionality of hazards or odds, are used routinely in statistical analysis. In quantile regression, very convenient assumptions are represented by the constant-slope model (e.g., \( \beta(u|\theta) = \theta_0 \)), in which a certain predictor has the same effect at all quantiles, and the linear-slope model (e.g., \( \beta(u|\theta) = \theta_0 + \theta_1 u \)), in which a quantile regression coefficient is assumed to be a linear function.

### 4. The Estimator

Frumenuto and Bottai (2016) considered cross-sectional data \((y_i, x_i)\) and defined \( \beta(U_i|\theta) = \theta b(u) \) as in (3). To estimate \( \theta \), they suggested minimizing

\[
L(\theta) = \int_0^1 \sum_i \rho_u(y_i - x_i^T \beta(U_i|\theta))du, \tag{5}
\]

which is the integral, with respect to the order of the quantile, of the loss function of standard quantile regression, \( \rho_u(w) = w(u - I(w \leq 0)) \) being the "check" function. This estimation method is referred to as integrated loss minimization (ILM) and is currently implemented in the \texttt{qcm} R package.

To generalize this idea to longitudinal data, assume model (4) holds,

\[
y_{it} = x_{it}^T \beta(U_{it}|\theta) + z_{it}^T \psi(V_i|\phi) = x_{it}^T \beta(U_{it}) + z_{it}^T \phi c(V_i),
\]

denote by \( y_{it} \) a realization of \( Y_{it} \). If the individual effects \( \alpha_i = z_i^T \phi c(V_i) \) were known, one could directly apply the ILM estimator to \( y_{it} - \alpha_i \), to compute an estimate of \( \theta \); and to \( \alpha_i \), to compute an estimate of \( \phi \). This would require solving

\[
\min_{\theta} L_1(\theta, \alpha_N), \min_{\phi} L_2(\phi, \alpha_N)
\]

where \( \alpha_N = (\alpha_1, \ldots, \alpha_N) \),

\[
L_1(\theta, \alpha_N) = \int_0^1 \sum_i \rho_u(y_{it} - \alpha_i - x_{it}^T \beta(U_{it}|\theta))du \tag{6}
\]

\[
= \sum_{i=1}^N \left\{ (y_{it} - \alpha_i)(u_{it}(\theta, \alpha_i) - 0.5) + x_{it}^T \theta [\bar{B} - B(u_{it}(\theta, \alpha_i))] \right\},
\]

\[
L_2(\phi, \alpha_N) = \int_0^1 \sum_i \rho_v(\alpha_i - z_{it}^T \psi(V_i|\phi))dv \tag{7}
\]

\[
= \sum_{i=1}^N \left\{ \alpha_i(v_i(\phi, \alpha_i) - 0.5) + z_{it}^T \phi [\bar{C} - C(v_i(\phi, \alpha_i))] \right\}.
\]

To obtain expressions (6)\(^2\) and (7), we used equation (9) from Frumento and Bottai (2016), and define

\[
B(u) = \int_0^u b(s)ds, \bar{B} = \int_0^1 B(u)du, \tag{8}
\]

\[
C(v) = \int_v^\infty c(s)ds, \bar{C} = \int_0^1 C(v)dv. \tag{9}
\]

In the formulas, \( u_{it}(\theta, \alpha_i) \) and \( v_i(\phi, \alpha_i) \) are such that \( y_{it} - \alpha_i = x_{it}^T \beta(U_{it}|\theta, \alpha_i) \) and \( \alpha_i = z_i^T \phi c(V_i, \phi) \), respectively. This also implies that

\[
u_{it}(\theta, \alpha_i) = F_{y_{it} - \alpha_i|x_{it}, \theta} \tag{10}
\]

is the cumulative distribution of \( Y_{it} - \alpha_i \), given \( x_{it} \), with parameter \( \theta \); and

\[
v_i(\phi, \alpha_i) = F_{c} (\alpha_i|z_i, \phi) \tag{11}
\]

is the cumulative distribution of \( \alpha_i \), given \( z_i \), with parameter \( \phi \).

In practice, the vector \( \alpha_N \) of individual effects is not known and must be estimated. We propose estimating \( (\theta, \phi, \alpha_N) \) by solving

\[
\min_{\theta, \phi, \alpha_N} L_1(\theta, \alpha_N) + L_2(\phi, \alpha_N). \tag{12}
\]

The proposed loss function is similar to that of a penalized fixed-effects estimator in which \( L_2(\phi, \alpha_N) \) plays the role of a penalty term. Intuitively, \( L_2(\phi, \alpha_N) \) shrinks the estimated fixed effects toward their assumed conditional distribution, introducing some degree of smoothing, improving model identification and efficiency, and avoiding overfitting. At the same time,

\[^2\] We index \( \alpha_N \) by \( N \) to emphasize that the dimension grows with the sample size.

\[^3\] The expression for \( L_1(\theta, \alpha_N) \) bears some similarity to Koenker’s (2004) loss function for unpenalized fixed-effects quantile regression, which is defined by \( L(\beta, \alpha_N) = \sum_i \sum_j w_{ij}(y_{ij} - \alpha_i - x_{ij}\beta(u_{ij})) \) and can be seen as a discretized, nonparameterized, and weighted version of \( L_1(\theta, \alpha_N) \).
\(L_2(\phi, \alpha_N)\) carries information on the parameter \(\phi\) that describes the quantile function of \(\alpha_N\).

Since both \(\alpha_N\) and \(\phi\) are treated as parameters, this approach combines features of fixed-effects estimators, which only estimate \(\theta\) and \(\alpha_N\), and random-effects models, which directly estimate \(\theta\) and \(\phi\). Computation, however, is much simpler than that of purely random-effects methods (e.g., Kim and Yang 2011; Arellano and Bonhomme 2016).

The gradient functions of \(L(\theta, \phi, \alpha_N) = L_1(\theta, \alpha_N) + L_2(\phi, \alpha_N)\) can be written as

\[
G_\theta(\theta, \alpha_N) = \nabla_{\text{vec}(\theta)} L(\theta, \phi, \alpha_N) = \sum_{i=1}^{N} \left[ B - B(u_i(\theta, \alpha_i)) \right] \otimes x_i, \tag{13}
\]

\[
G_\phi(\phi, \alpha_N) = \nabla_{\text{vec}(\phi)} L(\theta, \phi, \alpha_N) = \sum_{i=1}^{N} \left[ \bar{C} - C(v_i(\phi, \alpha_i)) \right] \otimes z_i, \tag{14}
\]

\[
G_{\alpha_i}(\alpha_i, \theta, \phi) = \nabla_{\alpha_i} L(\theta, \phi, \alpha_N) = \left[ \sum_{i=1}^{T} (0.5 - u_i(\theta, \alpha_i)) \right]^{\top} + (v_i(\phi, \alpha_i) - 0.5), \tag{15}
\]

where vec denotes the vectorization operator and \(\otimes\) the kronecker product. The model parameters, \((\theta, \phi, \alpha_N)\), only enter Equations (13)–(15) through the cumulative distribution functions \(u_i(\theta, \alpha_i)\) and \(v_i(\phi, \alpha_i)\) defined in (10) and (11). Note that \(G_\theta(\theta, \alpha_N)\) does not carry information on \(\phi\), and \(G_\phi(\phi, \alpha_N)\) does not carry information on \(\theta\); while \(G_{\alpha_i}(\alpha_i, \theta, \phi)\) depends on both \(\theta\) and \(\phi\). As shown by Frumento and Bottai (2016), \(G_\theta(\theta, \alpha_N)\) and \(G_\phi(\phi, \alpha_N)\) approach zero when the distributions of \(u_i(\theta, \alpha_i)\) and \(v_i(\phi, \alpha_i)\) tend to be uniform. This reflects the data-generating process described in (1), which involves the two independent uniform variables \(U_i\) and \(V_i\).

Equation (15) clarifies the role of the “penalty” term \(L_2(\phi, \alpha_N)\):

- the left-hand side of (15), \(\sum_{i=1}^{T} (0.5 - u_i(\theta, \alpha_i)) = \nabla_{\alpha_i} L_1(\theta, \alpha_N)\), is an unpenalized estimating equation for \(\alpha_i\). It approaches zero when \(u_1(\theta, \alpha_i), u_2(\theta, \alpha_i), \ldots, u_T(\theta, \alpha_i)\) are evenly spaced in \((0, 1)\), imposing a within-cluster uniformity of \(u_i(\theta, \alpha_i)\) which mirrors the assumption of independence between \(U_i\) and \(V_i\);
- the right-hand side, \(v_i(\phi, \alpha_i) - 0.5\), is a penalty term that shrinks the value of \(\alpha_i\) toward its conditional median, \(z_i^T \gamma(0.5) \phi = z_i^T \gamma(0.5)\).

A desirable property of the proposed penalization is that it only affects the estimates of \(\alpha_N\) when the clusters are relatively small. As \(T \rightarrow \infty\), each cluster contains sufficient information to estimate its own individual effect and, consistently, the penalty term \((v_i(\phi, \alpha_i) - 0.5)\) in Equation (15) becomes irrelevant.

Estimation can be performed by the following iterative process: (i) given \(\alpha_N\), estimate \(\theta\) and \(\phi\) separately by solving \(G_\theta(\theta, \alpha_N) = 0\) and \(G_\phi(\phi, \alpha_N) = 0\); (ii) given \((\theta, \phi)\), compute a new estimate of \(\alpha_N\) by solving \(G_{\alpha_i}(\alpha_i, \theta, \phi) = 0\), \(i = 1, \ldots, N\). Step (i) can be implemented with standard routines available in the \texttt{qzcm} package, while Step (ii) requires finding the zero of \(N\) univariate estimating equations. Neither \(u_i(\theta, \alpha_i)\) nor \(v_i(\phi, \alpha_i)\) are generally available in closed form, and can be evaluated by using a bijection algorithm. Note that the objective function defined by (12) is a smooth function of all parameters, unlike the loss function of standard quantile regression.

The fact that the quantile function may be ill-defined at some value of the parameters can be an issue during estimation. In the implementation of the \texttt{qzcm} package, we use unconstrained optimization from carefully chosen initial values. The algorithm is described in detail in Appendix B.

### 4.1. A New Family of Penalized Fixed-Effects Estimators

A possible interpretation of the proposed loss function,

\[
L(\theta, \phi, \alpha_N) = L_1(\theta, \alpha_N) + L_2(\phi, \alpha_N),
\]

is to consider \(L_2(\phi, \alpha_N)\) as a penalty term that shrinks the estimated individual effects toward their conditional median, \(z_i^T \gamma(0.5) \phi\). Unlike standard penalizations, however, \(L_2(\phi, \alpha_N)\) may depend on level-2 covariates and is a function of estimated parameters. To clarify this idea, consider a more traditional penalized loss function

\[
L_\lambda(\theta, \phi, \alpha_N) = L_1(\theta, \alpha_N) + \lambda L_2(\phi, \alpha_N),
\]

where \(L_2(\alpha_N)\) is a penalty term which does not contain \(\phi\), and \(\lambda\) is a tuning parameter. Common choices of \(L_2(\alpha_N)\) are the \(\ell_1\)-penalization, \(L_2(\alpha_N) = \sum_{i=1}^{N} |\alpha_i|\), which was used by Koenker (2004) to implement longitudinal quantile regression, and the \(\ell_2\)-penalization, \(L_2(\alpha_N) = \sum_{i=1}^{N} \alpha_i^2\).

Standard \(\ell_1\) and \(\ell_2\)-penalized fixed-effects methods are computationally simple and can substantially improve efficiency of the estimates of the structural parameters. However, besides the fact that they do not allow for estimation of \(\phi\), they present some important limitations: (i) they do not use prior knowledge on the distribution of the individual effects; (ii) they can introduce bias; (iii) they apply the same penalization to all clusters; and (iv) they require to specify a tuning parameter.

For instance, \(\ell_1\)-penalized estimators of quantile regression coefficients are asymptotically biased unless \((\alpha_1, \ldots, \alpha_N)\) are independent and identically distributed with zero median (Lamarche 2010). This is just a consequence of the \(\ell_1\)-penalty term being a sum of absolute deviations from zero, which does not generally reflect the true distribution of \(\alpha_N\) and the effect of level-2 covariates on it. Moreover, the same value of \(\lambda\) is used for all clusters, ignoring the fact that the variance of the individual effects may differ across subgroups of the population.

The tuning constant \(\lambda\) determines the degree of shrinking and, in the standard random-intercept linear model, its optimal value is \(\hat{\sigma}_e^2/\hat{\sigma}_\alpha^2\), that is, a function of nuisance scale parameters (e.g., Koenker 2004). Outside the restrictive conditions of linear models, not only the choice of \(\lambda\) becomes problematic, but also the use of a single value of \(\lambda\) for all clusters is questionable.

The novelty of our approach is that, unlike the \(\ell_1\)- and \(\ell_2\)-penalizations, the term \(L_2(\phi, \alpha_N)\) reflects the true (conditional) distribution of \(\alpha_N\) and carries information about its parameters,
Our estimator presents the following advantages over standard penalized methods: (i) it enables incorporating parametric assumptions on the distribution of $x_{it}$; (ii) it permits estimating all parameters consistently; (iii) it applies a different degree of shrinking to each cluster, by modeling the effect of level-2 covariates on the distribution of the individual effects; and (iv) it does not require selecting a tuning constant, as no nuisance parameters are present.

To clarify point (iv), consider the loss function of an $L_2$-penalized linear regression model: $L_1(\theta, \phi, \alpha_N) = L_1(\theta, \phi, \alpha_N) + \lambda_2(\alpha_N) = \sum_{t=1}^{N} \sum_{i=1}^{T} (y_{it} - x_{it}^T \beta - \alpha_i)^2 + \lambda \sum_{i=1}^{N} \alpha_i^2$. Here, $L_1(\theta, \phi, \alpha_N)$ and $L_2(\phi, \alpha_N)$ lack information on the nuisance scale parameters $\sigma^2_\epsilon = \text{var}(\epsilon_{it})$ and $\sigma^2_\alpha = \text{var}(\alpha_i)$. This is adjusted for by the tuning constant $\lambda = \sigma^2_\epsilon/\sigma^2_\alpha$. In our special type of penalized estimator, instead, $L_1(\theta, \phi, \alpha_N)$ and $L_2(\phi, \alpha_N)$ carry information on all model parameters. Intuitively, this means that $L_1(\theta, \phi, \alpha_N)$ and $L_2(\phi, \alpha_N)$ are already “properly scaled.” The tuning constant can be thought of as an implicit parameter, a function of $\theta$ and $\phi$. Although a more general estimator with criterion function $L_1(\theta, \phi, \alpha_N) + \lambda_2(\phi, \alpha_N)$ could in principle be formulated, choosing $\lambda = 1$ appears natural and avoids the problem of selecting the tuning parameter.

5. Inference

The asymptotic properties of fixed-effects estimators are complicated by the fact that, as $N \to \infty$, the dimension of the parameter $\alpha_N$ tends to infinity. Unless $T \to \infty$, the individual effects $\alpha_i$ are estimated using a fixed number of observations. This is often referred to as the “incidental parameter” problem (Neyman and Scott 1948; Lancaster 2000), which causes widely used estimators, such as maximum likelihood and $M$-estimators, to be inconsistent.

To develop the asymptotic theory of our estimator, we follow the recent panel data literature in econometrics and deal with the incidental parameter problem by considering asymptotic sequences where both $N$ and $T$ tend to infinity (e.g., Phillips and Moon 1999; Hahn and Newey 2004; Koenker 2004; Fernández-Val 2005; Arellano and Hahn 2007; Lamarche 2010; Kato, Galvao, and Montes-Rojas 2012). Under this approximation, we show that our estimators are consistent but might have biases in the asymptotic distribution depending on the relative rate of convergence of $N$ and $T$. We apply the theory of $M$-estimators (e.g., Newey and McFadden 1994), and use well-established results to handle the following nonstandard features of our problem: (i) the estimators of $\theta$, $\phi$, and $\alpha_N$ converge at different rates (e.g., Radchenko 2008; Cheng and Wang 2015; Masuda and Shimizu 2017); and (ii) additional conditions are required on the relative growth rate of $N$ and $T$ (e.g., Hahn and Newey 2004; Fernández-Val 2005; Newey 2007).

Let $x_{it} = (x_{i1}^T, x_{it2}^T)^T$, where $x_{i1}$ contains the time-invariant components including the constant and $x_{it2}$ contains the time-varying covariates. We use the following sufficient conditions to establish the identification of the parameters and derive the asymptotic properties of the estimators:

**Assumption 1 (Longitudinal ILM estimator).** (i) The data generating process is $Y_{it} = \theta^T x_{it}^T + z_{it}^T \phi^T c(V_i)$, where $Y_{it} = \theta^T x_{it}^T + z_{it}^T \phi^T c(V_i)$, $\phi^T c(V_i)$, and $\phi^T c(V_i)$ are positive definite. (ii) There exist two sets of quantile indexes $[u_1, \ldots, u_{kd}]$ and $[v_1, \ldots, v_{kd}]$ such that the matrices $b(u_1, \ldots, u_{kd}) := [b(u_1), \ldots, b(u_{kd})]$ and $c(v_1, \ldots, v_{kd}) := [c(v_1), \ldots, c(v_{kd})]$ have full rank. (v) The functions $u \mapsto b(u)$ and $v \mapsto c(v)$ are three times continuously differentiable on $(0, 1)$, sup $|b(u)|, |c(v)| < \infty$ for some $\xi > 32$, and $||x_{it}^T z_{it}^T||$ is bounded a.s. (vi) The following probability limits exist:

$$
\begin{align*}
\hat{H}_\theta &= \text{plim}_{N,T \to \infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \text{E}(H_{\theta it}), \\
\hat{H}_\phi &= \text{plim}_{N,T \to \infty} \frac{1}{N} \sum_{i=1}^{N} \text{E} \left( \frac{c(V_i) c(V_i)^T}{z_i^T \phi^T c(V_i)} \otimes z_i z_i^T \right), \\
\hat{H}_{\phi \theta} &= \text{plim}_{N,T \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \text{E} \left( b(U_{it}) \otimes x_{it} \right), \\
\tilde{b}_\theta &= \text{plim}_{N,T \to \infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \text{E}(b_{\theta it}), \\
\tilde{b}_\phi &= \text{plim}_{N,T \to \infty} \frac{1}{N} \sum_{i=1}^{N} \text{E}(b_{\phi it}), \\
\hat{S}_\theta &= \text{plim}_{N,T \to \infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \text{E}(\phi_{\theta it} \phi_{\theta it}), \\
\hat{S}_\phi &= \text{plim}_{N,T \to \infty} \frac{1}{N} \sum_{i=1}^{N} \text{E} \left( \left[ C - C(V_i) \right] \left[ C - C(V_i) \right]^T \otimes z_i z_i^T \right),
\end{align*}
$$

where the expectation $E$ is taken with respect to the distribution of $U_{it}$ and $V_i$, $C$ denotes the Kronecker product, and $c'$ and $c''$ denote the vectors of first and second derivatives of $v \mapsto c(v)$, respectively.
It can be replaced by any other existence and uniqueness conditions, together with a location normalization on the fixed effects. The distribution of the estimators in large samples. The derivatives of all the model parameters. Assumption 1(v) imposes that the conditional quantile and density functions (6) and (7) and their partial derivatives, which are needed to carry out higher-order expansions of these functions, and (vii) impose regularity conditions to derive the distribution of the estimators in large samples. The large sample distribution of the plugin estimators of and can be obtained by the delta method. Let (u) → θ(u) and (v) = φ(c(v)), for any u, v ∈ (0, 1). Then, if N = O(T),

\[
\sqrt{N} \left( \frac{\hat{\phi} - \phi^0}{T} + \frac{\hat{H}_\phi^{-1/2} \hat{H}_\phi^{-1} \hat{b}_\phi}{T} \right) \rightarrow_d \hat{H}_\phi^{-1} N(0, \tilde{\Phi}_\phi).
\]

The expressions of all the terms are given in Assumption 1.

Theorem 1 shows that the parameters θ0 and φ0 are identified and their estimators and have a normal distribution in large samples with different rates of convergence. The large sample distribution of the plugin estimators of and can be obtained by the delta method. Let (u) → θ(u) and (v) = φ(c(v)), for any u, v ∈ (0, 1). Then, if N = O(T),

\[
\sqrt{NT} \left( \frac{\hat{\theta} - \theta^0}{T} + \frac{\hat{H}_\theta^{-1/2} \hat{H}_\theta^{-1} \hat{b}_\theta}{T} \right) \rightarrow_d \hat{H}_\theta^{-1} N(0, \tilde{\Phi}_\theta).
\]

The rates of convergence of all the estimators agree with the square roots of the dimensions of the observations that are informative about the corresponding parameters. Thus, the rate is \(\sqrt{NT}\) for θ0 and \(\theta^0(u)\), and \(\sqrt{N}\) for \(\phi^0\) and \(\phi^0(u)\). All the estimators might suffer from bias in short panels due to the estimation of the fixed effects. The order of this bias is the inverse of the number of observations that are informative about each fixed effect, that is, \(T^{-1}\). Comparing the rates of convergence with the order of the bias, we can see that the biases of \(\hat{\theta}\) and \(\hat{b}(u)\) are negligible in the asymptotic distribution when \(N/T \rightarrow 0\), whereas the biases of \(\hat{\phi}\) and \(\hat{\gamma}(v)\) are negligible when \(N/T^2 \rightarrow 0\). These biases can be reduced by using analytical or jackknife corrections (e.g., Hahn and Newey 2004; Fernández-Val 2005; Dhaene and Jochmans 2015). We provide consistent analytical estimators of the components of the biases and variances below.

We construct estimators of the components of the asymptotic distribution using sample analogs evaluated at the estimated

We normalize the mean of the fixed effects. Alternative normalizations on the median or other quantile of the fixed effects are also possible.
value of the parameters, for example, \( \hat{u}_{it} = u_{it}(\hat{\theta}, \hat{\alpha}_i) \) and \( \hat{v}_i = v_i(\hat{\phi}, \hat{\alpha}_i) \). Then,

\[
\hat{H}_{\theta} = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{H}_{\theta(it)}, \quad \hat{H}_{\phi} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{c(\hat{v}_i)c(\hat{v}_i)}{\hat{z}_i^0 \hat{\phi}(c(\hat{v}_i))} \otimes \hat{z}_i^2 \right) T^{-1},
\]

\[
\hat{b}_{\theta} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{b}_{\theta(it)}, \quad \hat{b}_{\phi} = \frac{1}{N} \sum_{i=1}^{N} \hat{b}_{\phi(i)};
\]

where

\[
\hat{H}_{\theta(it)} = \frac{\hat{b}(\hat{u}_{it}) b(\hat{u}_{it})}{x_i^0 \hat{\theta}(\hat{u}_{it})} \otimes x_{it} - \frac{\hat{b}(\hat{u}_{it})}{x_i^0 \hat{\theta}(\hat{u}_{it})} \otimes x_{it} \hat{\sigma}_i^2 T^{-1},
\]

\[
\hat{H}_{\phi(it)} = \frac{c(\hat{v}_i)c(\hat{v}_i)}{\hat{z}_i^0 \hat{\phi}(c(\hat{v}_i))} \otimes z_i \hat{\sigma}_i^2 T^{-1} - \frac{c(\hat{v}_i)}{\hat{z}_i^0 \hat{\phi}(c(\hat{v}_i))} \otimes z_i \hat{\sigma}_i^2 T^{-1} \times \left( \frac{c(\hat{v}_i)}{\hat{z}_i^0 \hat{\phi}(c(\hat{v}_i))} \otimes z_i \right),
\]

• Step 0: Compute a test statistic \( D \) that measures the distance between \( F_{\hat{u}_{it}}(\hat{u}_{it}, \hat{v}_i) \) and \( F_{u_{it}}(\hat{u}_{it}, \hat{v}_i) = \tilde{u}_{it} \tilde{v}_i \).

• Step 1: Simulate new data as \( Y^s_{it} = x_i^0 \hat{\theta}(U^s_{it}) + \hat{z}_i^0 \hat{\phi}(V^s_{it}) \) by randomly generating \( (U^s_{it}, V^s_{it}) \) from two independent \( U(0, 1) \) distributions.

• Step 2: Fit the model on the simulated data and compute the corresponding value \( D^s \) of the test statistic.

After repeating Steps 1 and 2 for a sufficient number of times, the \( p \)-value is computed as the empirical proportion of cases in which \( D^s > D \). In Step 1, it is also possible to take a random sample of clusters, and to resample the covariates’ value within each cluster. To assess local fit, the test could be repeated within subsets of the original sample identified by specific values of the covariates.

In the implementation of the \texttt{QFMC} package, we chose \( D \) to be the Kolmogorov–Smirnov statistic, \( \sup_{u,v} |F_{\hat{u}_{it}}(u, v) - F_{u_{it}}(u, v)| \). This testing procedure is usually reliable, as indicated by the simulation results reported in Section 7.

### 6.2. Model Selection

As suggested in Section 3.5 and exemplified in the real-data example presented in Section 8, it is usually possible to identify numerous alternative models that have a similar fit and are not rejected by a goodness-of-fit test. This can be explained by the fact that the same coefficient functions can be well approximated by different parametric functions.

Important criteria for model selection include parsimony, flexibility, and interpretability. Nested models can be compared by standard Wald test. Let \( \hat{\alpha}_N = (\hat{\alpha}_1, \ldots, \hat{\alpha}_N) \). To compare nonnested models, the value of \( L_1(\hat{\theta}, \hat{\alpha}_N) \) and \( L_2(\hat{\phi}, \hat{\alpha}_N) \) can be used to construct information criteria such as the AIC (Akaike 1974) and the BIC (Schwarz 1978). These criteria were initially designed for likelihood-based estimators, but can be extended to estimators defined by the minimizer of a loss function. For example, a modification of BIC criterion for \( M \)-estimators has been described by Machado (1993), while Koeken (2005) used AIC to compare quantile regression models. Consider the probability density function of the asymmetric Laplace distribution,

\[
f_p(y|\mu, \sigma) = \frac{p(1-p)}{\sigma(p)} \exp \left\{ -\frac{\rho(y - \mu(p))}{2\sigma(p)} \right\},
\]
where $\mu(p)$ is a location parameter and corresponds to the $p$th quantile of the distribution, while $\sigma(p)$ is a scale parameter. Although this distribution is not generally considered a plausible model, its log-likelihood has been used by numerous authors, including Koenker and Machado (1999) and Lee, Noh, and Park (2014), to obtain measures of goodness of fit for quantile regression. Simple algebra permits showing that

$$
\hat{\theta} = \arg \min_{\theta} \int_0^1 \sum_{i=1}^N \sum_{t=1}^T \log f_{\mu}(y_{it} - \hat{\alpha}_i | x_i^T \theta b(u), \sigma_1) du,
$$

$$
\hat{\phi} = \arg \min_{\phi} \int_0^1 \sum_{i=1}^N \sum_{t=1}^T \log f_{\mu}(\hat{\alpha}_i | x_i^T \phi(c(v), \sigma_2)) dv,
$$

that is, $\hat{\theta}$ and $\hat{\phi}$ minimize an "average" Laplace log-likelihood, in which $u$ and $v$ have been integrated away. After substituting $\hat{\alpha}_1 = L_1(\hat{\theta}, \hat{\alpha}_N)/(2NT)$ and $\hat{\alpha}_2 = L_2(\hat{\phi}, \hat{\alpha}_N)/(2N)$, we obtain the following AIC and BIC:

$$
\text{AIC}_1 = \log L_1(\hat{\theta}, \hat{\alpha}_N) + \frac{q_1}{NT}, \quad \text{BIC}_1 = \log L_1(\hat{\theta}, \hat{\alpha}_N) + \frac{q_1 \log(NT)}{2NT},
$$

$$
\text{AIC}_2 = \log L_2(\hat{\phi}, \hat{\alpha}_N) + \frac{q_2}{N}, \quad \text{BIC}_2 = \log L_2(\hat{\phi}, \hat{\alpha}_N) + \frac{q_2 \log(N)}{2N},
$$

where $q_1$ and $q_2$ are the number of nonzero elements of $\theta$ and $\phi$, respectively. Note that $\text{BIC}_1$ can be obtained from equation 2.3 of Lee, Noh, and Park (2014) by replacing the loss of standard quantile regression with $L_1(\theta, \hat{\alpha}_N)$.

The proposed criteria seem to work well in simulation (see Appendix C). However, they often tend to reward parsimony, possibly sacrificing goodness of fit. The testing procedure described in Section 6.1 should always be used to perform a preliminary screening of the candidate models.

### 7. Simulation Results

We analyze the performance of the estimators $\hat{\beta}(u)$ and $\hat{\gamma}(v)$ in finite samples through numerical simulations. In particular, we report the biases and standard errors of these estimators for different values of the dimensions $T$ and $N$ and the orders of the quantiles $u$ and $v$. We also evaluate the empirical size and power of the goodness-of-fit test.

We used the following design to generate the data:

$$
Y_{it} = \beta_0(U_{it}) + \beta_1(U_{it})x_{it} + \gamma_0(V_i) + \gamma_1(V_i)z_i,
$$

where $x_{it} \sim \text{Beta}(2,2)$ and $z_i \sim \text{U}(0,3)$. In simulation 1, we defined:

$$
\beta_0(u) = 1 - 0.5 \log(1 - u), \quad \beta_1(u) = 1 + 10(u - 0.5)^3,
$$

$$
\gamma_0(v) = 0.5 \zeta(v), \quad \gamma_1(v) = 0.5 \zeta(v),
$$

where $\zeta(v)$ is the quantile function of a standard normal distribution. In simulation 2, we defined:

$$
\beta_0(u) = 2(1 - (1 - u)^{1/4}), \quad \beta_1(u) = 3(1 + u),
$$

$$
\gamma_0(v) = \log(1 - \log(1 - v)), \quad \gamma_1(v) = 0.5 \log(1 - \log(1 - v)).
$$

### Table 1. Simulation results with $N = 150$.

| $T = 5$ | Simulation 1 | Simulation 2 |
|---------|--------------|--------------|
| $u$     | $\beta_0$    | $\hat{\beta}_0$ | $SE$ | $\hat{\beta}_1$ | $SE$ | $\hat{\beta}_1$ | $SE$ |
| 0.2     | 1.11         | 1.17         | 0.10 | 0.11            | 0.73 | 0.72         | 0.10 | 0.11            |
| 0.4     | 1.26         | 1.30         | 0.11 | 0.11            | 0.99 | 0.97         | 0.11 | 0.11            |
| 0.6     | 1.46         | 1.48         | 0.11 | 0.11            | 1.01 | 0.99         | 0.11 | 0.11            |
| 0.8     | 1.80         | 1.79         | 0.12 | 0.12            | 1.27 | 1.24         | 0.12 | 0.13            |
| $\nu$   | $\gamma_0$   | $\gamma_0$   | $SE$ | $\gamma_1$     | $\gamma_1$ | $SE$ | $\gamma_1$     | $\gamma_1$ | $SE$ |
| 0.2     | 0.42         | 0.37         | 0.09 | 0.10            | 0.42 | 0.40         | 0.08 | 0.08            |
| 0.4     | 0.13         | 0.11         | 0.03 | 0.03            | 0.13 | 0.12         | 0.02 | 0.02            |
| 0.6     | 0.13         | 0.11         | 0.03 | 0.03            | 0.13 | 0.12         | 0.02 | 0.02            |
| 0.8     | 0.42         | 0.37         | 0.09 | 0.10            | 0.42 | 0.40         | 0.08 | 0.08            |
| $T = 10$ | $\beta_0$    | $\hat{\beta}_0$ | $SE$ | $\hat{\beta}_1$ | $SE$ | $\hat{\beta}_1$ | $SE$ |
| $u$     | $\beta_0$    | $\hat{\beta}_0$ | $SE$ | $\hat{\beta}_1$ | $SE$ | $\hat{\beta}_1$ | $SE$ |
| 0.2     | 1.11         | 1.14         | 0.10 | 0.10            | 0.73 | 0.73         | 0.06 | 0.07            |
| 0.4     | 1.26         | 1.27         | 0.10 | 0.10            | 0.99 | 0.98         | 0.07 | 0.07            |
| 0.6     | 1.46         | 1.47         | 0.10 | 0.10            | 1.01 | 1.00         | 0.07 | 0.07            |
| 0.8     | 1.80         | 1.80         | 0.10 | 0.10            | 1.27 | 1.25         | 0.08 | 0.08            |
| $\nu$   | $\gamma_0$   | $\gamma_0$   | $SE$ | $\gamma_1$     | $\gamma_1$ | $SE$ | $\gamma_1$     | $\gamma_1$ | $SE$ |
| 0.2     | 0.42         | 0.40         | 0.09 | 0.09            | 0.42 | 0.41         | 0.08 | 0.08            |
| 0.4     | 0.13         | 0.12         | 0.03 | 0.03            | 0.13 | 0.12         | 0.02 | 0.02            |
| 0.6     | 0.13         | 0.12         | 0.03 | 0.03            | 0.13 | 0.12         | 0.02 | 0.02            |
| 0.8     | 0.42         | 0.40         | 0.09 | 0.09            | 0.42 | 0.41         | 0.08 | 0.08            |

### Note

Summary of simulation results, based on $R = 1000$ Monte Carlo replications, with $N = 150$ and $T = \{5, 10\}$. For each coefficient, we report the true absolute value ($\beta, \gamma$) at the quantiles ($0.2, 0.4, 0.6, 0.8$), the average estimate ($\hat{\beta}, \hat{\gamma}$), the standard error (SE), computed as the standard deviation of the estimated model parameters across simulations, and the average estimated asymptotic standard error (SE). The bottom table reports, for two different nominal levels $\alpha = 0.05, 0.10$, the empirical probability of Type I error ($\hat{\alpha}$) and the power ($1 - \hat{\beta}$) of the Kolmogorov–Smirnov goodness-of-fit test described in Section 6.
To fit the true model, we used $b(u) = [1, -\log(1 - u), (u - 0.5)^2]$ and $c(v) = [\gamma(v)]$ in simulation 1, and $b(u) = [1, 1 / \sqrt{1 - u}, u^T]$ and $c(v) = [\log(1 - \log(1 - v))]$ in simulation 2. We ran $R = 1000$ Monte Carlo simulations, with $N = \{150, 300\}$ and $T = \{5, 10\}$. In Tables 1 and 2, we report the true value of $\beta(\cdot)$ and $\gamma(\cdot)$ at the quintiles, their average estimates, the empirical standard errors across simulations, and the average estimates of the asymptotic standard errors. Despite the incidental parameters problem, a small bias was found, even with small values of $T$. Also, as $T$ increased, the observed bias decreased rapidly as predicted by the asymptotic theory of Section 5. The estimated standard errors were, on average, very close to their true values.

To assess the performance of the goodness-of-fit procedure described in Section 6.1, we selected two nominal significance levels, $\alpha = 0.05$ and $\alpha = 0.10$, and computed the empirical probability of Type I error ($\hat{\alpha}$) and the power ($1 - \hat{\beta}$) of the Kolmogorov–Smirnov goodness-of-fit test described in Section 6. The power was estimated by the empirical probability to reject a misspecified model in which the quantile function was described by an incorrect basis function. In simulation 1, we incorrectly parameterized $\beta_1(u)$ as a linear function, $\beta_1(u) = \theta_{01} + \theta_{11}u$. In simulation 2, we incorrectly assumed that the individual effects have a logistic distribution, defined by $c(v) = \log(v/(1 - v)))$. Results are shown in the bottom rows of Tables 1 and 2. The risk of Type I error was very close to its nominal level, and approached it as the value of $T$ increased. With small values of $N$ and $T$, the risk of Type II error was relatively large, and the power was often less than 50%. However, with $N = 300$ and $T = 10$, and a nominal level of 0.10 for rejection, the incorrect models were rejected in more than 90% of cases in both scenarios.

Additional simulation results are reported in Appendix C, where we compare our estimator with Koenker’s (2004) penalized fixed-effects quantile regression, and discuss the performance of the model selection criteria presented in Section 6.2.

8. Analysis of NGAL Data

We analyzed data from Mårtensson et al. (2013), aiming to investigate the role of plasma NGAL as a marker of sepsis and acute kidney disfunction. The dataset included 139 patients admitted to the general intensive care unit at Karolinska University Hospital in Solna, Sweden, between August 2007 and November 2010. Baseline information was collected, and patients were classified daily as having sepsis or not. NGAL (mg/mL), procalcitonin (PCT), C-reactive protein (CRP), and creatinine changes relative to baseline (Acreat) were measured daily before discharge, for a total of 1317 plasma samples. After removing missing data, individuals with only one observation, and one patient with severe complications, the final sample included 135 patients for a total sample size of $\sum_{i=1}^{135} T_i = 1263$. The number of observations per patient varied between 2 and 38, and more than 80% of patients had $T_i \leq 14$.

The goal of our analysis was to estimate conditional quantiles of NGAL, and in particular to measure its association with sepsis. The between-patient variability appeared to be very large, reflecting the presence of important differences in the initial health conditions. We formulated a regression model with the following predictors: a binary indicator of sepsis status, an indicator of $\Delta$creat $\geq 50$, age (centered at its median, 52 years, and divided by 10), an indicator of female gender, and time since hospitalization (weeks). Age and gender were cluster-invariant and were also included as level-2 predictors.
Table 3. Alternative specifications of \( b(u) \) and \( c(v) \).

| \( \beta_0(u) \) | \( \beta_1(u) \) | \( \gamma_0(v) \) | \( \gamma_1(v) \) |
|-----------------|-----------------|-----------------|-----------------|
| \( \xi(u) = \log(\frac{u}{u_0}) \) | \( u, u^2, u^3, \ldots \) | \( \log(\frac{v}{v_0}) \) | \( (v - 0.5) \) |
| \( \log(u), \log(1 - u) \) | \( u, \cos(\pi u), \sin(\pi u) \) | \( \log(2v), \log(2(1 - v)) \) | \( (v - 0.5), (v - 0.5)^3 \) |

NOTE: In the table, \( \beta_0(u) \) and \( \gamma_0(v) \) denote level-1 and level-2 intercept, respectively, while \( \beta_1(u) \) and \( \gamma_1(v) \) represent coefficients associated with generic level-1 and level-2 covariates. Different parametric models are represented by the functions that compose \( b(u) \) and \( c(v) \). A constant term \( \beta_0(u) = 1 \) was always included. The notation \( \xi(\cdot) \) is used for the quantile function of a standard normal distribution, and we defined \( (v - 0.5)^3 = \gamma(v < 0.5) \).

Table 4. Estimated model parameters.

| Level 1 (\( \hat{\theta} \)) | | | |
|-----------------|-----------------|-----------------|-----------------|
| Intercept | 4.67 (0.07) | 0.12 (0.03) | -0.13 (0.03) |
| Sepsis | 0.39 (0.26) | - | -20.20 (0.23) |
| Age \((\geq 52)/10 \) | 0.00 (0.09) | - | 0.02 (0.07) |
| Female gender | 0.16 (0.31) | - | -0.40 (0.23) |
| Time (weeks) | -0.14 (0.09) | - | 0.12 (0.08) |

| Level 2 (\( \hat{\phi} \)) | | | |
|-----------------|-----------------|-----------------|-----------------|
| Intercept | 0.33 (0.08) | -0.29 (0.08) | - |
| Age \((\geq 52)/10 \) | - | - | 0.11 (0.07) |
| Female gender | - | - | -0.20 (0.28) |

NOTE: Summary of the selected model (top: \( \hat{\theta} \); bottom: \( \hat{\phi} \)), with estimated standard errors in brackets. The model is represented graphically in Figure 2, and selected quantiles are summarized in Table 5.

The response variable was log-transformed, which made it more plausible to define individual effects on the additive scale as in model (1). The regression function was

\[
\log(\text{NGAL}_{it}) = \beta_0(U_{it}) + \beta_1(U_{it})I(\text{Sepsis}_{it}) + \beta_2(U_{it})I(\text{Age} \geq 52)/10 + \beta_3(U_{it})I(\text{Female}_{it}) + \beta_5(U_{it})I(\text{Time}_{it}) + \gamma_0(V_{i}) + \gamma_1(V_{i})I(\text{Age} \geq 52)/10 + \gamma_2(V_{i})I(\text{Female}_{it}).
\]

We formulated a variety of models, in which \( \beta_0(\cdot) \) and \( \gamma_0(\cdot) \) were unbounded, while the other coefficients were modeled by bounded functions. To facilitate interpretation, we forced \( \gamma(0.5) = 0 \), assigning the individual effects a zero-median distribution in which level-2 covariates only affect the scale parameter.

Selected modeling options are illustrated in Table 3. Different models appeared to fit the data well, and were not rejected by the goodness-of-fit test described in Section 6. The following model combined simplicity and flexibility, and was selected for illustrative purposes:

\[
\beta_0(u|\theta) = \theta_0 + \theta_1 \log(u) + \theta_2 \log(1 - u),
\]

\[
\beta_1(u|\theta) = \theta_3 + \theta_{4j} u^{1/4} + \theta_{5j} (1 - u)^{1/4}, j = 1, \ldots, 5,
\]

\[
\gamma_0(v|\phi) = \phi_0 \log(2v) + \phi_2 \log(2(1 - v)),
\]

\[
\gamma_1(v|\phi) = \phi_3 (v - 0.5), j = 1, \ldots, 2.
\]

The level-1 and level-2 intercepts were described by different versions of the asymmetric logistic distribution. The coefficient functions associated with level-1 covariates were a combination of linear and root-4 functions, while those of the level-2 predictors were assumed to be linear.

Table 5. Summary of selected quantiles.

| Quantile | 0.2 | 0.4 | 0.6 | 0.8 |
|---------|-----|-----|-----|-----|
| Intercept | 4.51 (0.05) | 4.63 (0.05) | 4.73 (0.05) | 4.85 (0.05) |
| Sepsis | 0.12 (0.02) | 0.12 (0.02) | 0.12 (0.02) | 0.13 (0.03) |
| Age \((\geq 52)/10 \) | 0.10 (0.05) | 0.11 (0.04) | 0.10 (0.04) | 0.11 (0.05) |
| Female gender | 0.03 (0.07) | 0.01 (0.07) | -0.02 (0.07) | -0.04 (0.07) |
| Time (weeks) | -0.09 (0.01) | -0.09 (0.01) | -0.08 (0.01) | -0.08 (0.01) |

| Intercept | -0.42 (0.06) | -0.12 (0.01) | 0.12 (0.01) | 0.41 (0.05) |
| Age \((\geq 52)/10 \) | -0.03 (0.02) | -0.01 (0.01) | 0.01 (0.01) | 0.03 (0.02) |
| Female gender | 0.05 (0.08) | 0.02 (0.03) | -0.02 (0.03) | -0.05 (0.08) |

NOTE: Estimated regression coefficients at quantiles \((0.2, 0.4, 0.6, 0.8)\). Top table: \( \hat{\beta}(u|\theta) = \beta(u|\theta) = \beta(u); \) bottom table: \( \hat{\gamma}(v|\phi) = \gamma(v|\phi) = \phi(v). \) Estimated standard errors in brackets.

The \( p \)-value of the Kolmogorov–Smirnov test was 0.21. To assess local fit, the test was repeated in subsamples with different values of the covariates (e.g., the females, those with \( \Delta\text{Creat} > 50 \), etc.). No significant evidence of model misspecification was found.

All 27 model parameters are reported in Table 4, while regression coefficients at selected quantiles are summarized in Table 5. We represent graphically the quantile regression coefficient functions in Figure 2, where we also report a “nonparametric” fit obtained by modeling all coefficients as piecewise linear functions with knots at the deciles.

Results showed that the distribution of the individual effects was almost symmetric (as suggested by the fact that \( \phi_{01} \approx -\phi_{03} \)) and that its variance was not significantly affected by cluster-level predictors. Instead, all predictors apart from gender appeared to be associated with the level-1 response. In particular, the coefficients associated with Sepsis, \( \Delta\text{Creat} > 50 \) and age were consistently positive at all quantiles, while the coefficient of time was always negative. The sepsis status was associated with...
a percentile difference of about 0.12 at quantiles 0.2, 0.4, 0.6, 0.8. As shown by Figure 2, an even larger percentile difference was found at quantiles above 0.8.

9. Conclusions

We introduced a general framework for longitudinal quantile regression, extending the work of Frumento and Bottai (2016, 2017) on quantile regression coefficients modeling. We defined a two-level quantile function in which both the “within” and the “between” part of the distribution are described by a quantile regression model. This allows to investigate how covariates affect not only the level-1 response, but also the distribution of the individual effects, which is generally overlooked in the existing literature on longitudinal quantile regression. Identification is achieved by modeling the coefficient functions parametrically, and estimation is carried out by minimizing a smooth objective function.

The proposed method is computationally simple and can be viewed as a special type of penalized fixed-effects estimator that presents important elements of novelty of its own. The penalty term carries information on the parameters that describe the conditional distribution of the individual effects. This permits estimating both level-1 and level-2 parameters, as in random-effects models, but allows carrying out estimation and inference using fixed-effects techniques. Moreover, it avoids the problem of choosing a tuning constant as in standard \( \ell_1 \)- or \( \ell_2 \)-penalization. The described form of penalized fixed-effects

Figure 2. Continuous lines represent the estimated quantile regression coefficient functions, based on the parametric model summarized in Table 4. Shaded areas represents pointwise 95% confidence intervals. The dashed lines are obtained from a "nonparametric" model in which \( \beta(u) \) and \( \gamma(v) \) were fitted by piecewise linear functions with knots at the deciles. The dotted horizontal lines indicate the zero.
method is not limited to a quantile regression framework and could be applied to different estimation problems.

The proposed modeling framework can be generalized in different directions. An interesting possibility is to include multiple individual effects as in random-slope models. In our framework, individual effects are represented by a pure location shift as in Koenker (2004), Geraci and Bottai (2007, 2014), and Canay (2011). Using the proposed penalized fixed-effects approach, it is relatively simple to incorporate not only an individual intercept, $\alpha_i$, but also a set of individual slopes, say $\{\delta_{i1}, \delta_{i2}, \ldots\}$. This, however, would typically result in cumbersome computation and, unless $N$ and $T$ are sufficiently large, would probably undermine model identifiability. Using Koenker’s (2004) words: “At best we may be able to estimate an individual specific location-shift effect, and even this may strain credulity.”

Another interesting extension is represented by varying-coefficient models (e.g., Hastie and Tibshirani 1993; Fan and Zhang 1999, 2000; Chiang, Rice, and Wu 2001; Kim 2007) that could be implemented by allowing the level-1 regression coefficients to be functions of time. A possible approach is to describe the coefficients, say $\beta(u, t)$, using tensor products of splines. Finally, the proposed method could be used to estimate static and dynamic quantile autoregressive models (e.g., Arellano and Bonhomme 2016).

An important problem that has not been discussed in the article is represented by quantile crossing, occurring when either $x_i^T \beta(u|\theta) < 0$ or $z_i^T \gamma(v|\phi) < 0$. One may want to determine in advance which values of the parameters $\theta$ and $\phi$ would ensure that no crossing occurs, that is, that $x_i^T \beta(u|\theta)$ and $z_i^T \gamma(v|\phi)$ are monotonically increasing functions. This is only possible in very simple models with few covariates, or in presence of restrictive assumptions. However, simulation evidence suggests that parametric models are relatively immune to quantile crossing, compared with the “nonparametric” approaches based on ordinary quantile regression. Additionally, the parametric structure makes it particularly simple to verify crossing, taking advantage of the closed-form analytical expression of the quantile function, and admits the application of monotonization methods such as the rearrangement of Chernozhukov, Fernández-Val, and Galichon (2010) to produce increasing estimates of conditional quantiles.

This article is accompanied by an R package qrcm (Frumento 2021), which includes a function named $\text{iqrL}$ that performs model fitting, and a variety of auxiliary functions for prediction, plotting, and goodness-of-fit assessment. The documentation contains a rich set of examples, and can serve as tutorial for the practitioners. The package is available upon request to the authors.

Supplementary Materials

Version 3.0 of the qrcm R-package (to be installed as local zip file):
New package version including the $\text{iqrL}$ function that implements the estimator.

NGAL dataset: Dataset and R script used in the illustration of the method in Section 8 (txt file).

Appendix A: Proof of Theorem 1.
Appendix B: Computation.
Appendix C: Extended simulation results.

References

Abrevaya, J., and Dahl, C. M. (2008), “The Effects of Birth Inputs on Birthweight,” Journal of Business & Economic Statistics, 26, 379–397. [783]
Alfo, M., Salvati, N., and Ranalli M. G. (2017), “Finite Mixtures of Quantile and M-Quantile Regression Models,” Statistics and Computing, 27, 547–570. [783]
Akaike, H. (1974), “A New Look at the Statistical Model Identification,” IEEE Transactions on Automatic Control, 19, 716–723. [791]
Angrist, J. D., Chernozhukov, V., and Fernandez-Val, I. (2006), “Quantile Regression Under Misspecification, With an Application to the U.S. Wage Structure,” Econometrica, 74, 539–563. [790]
Arelano, M., and Bonhomme, S. (2016), “Nonlinear Panel Data Estimation via Quantile Regression,” The Econometric Journal, 19, C61–C94, DOI: 10.1111/ectj.12062. [783,788,796]
Arelano, M., and Hahn, J. (2007), “Understanding Bias in Nonlinear Panel Models: Some Recent Developments,” Econometric Society Monographs, 43, 381. [789]
Canay, I. A. (2011), “A Simple Approach to Quantile Regression for Panel Data,” Econometrica, 14, 368–386. [783,796]
Chamberlain, G. (1984), “Panel Data” (Chapter 22), in Handbook of Econometrics, eds. Z. Griliches and M. Intrilligator, Amsterdam: Elsevier, pp. 1247–1318 [783]
Cheng, G., and Shang, Z. (2015), “Joint Asymptotics for Semi-Nonparametric Regression Models With Partially Linear Structure,” The Annals of Statistics, 43, 1351–1390. [789]
Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010), “Quantile and Probability Curves Without Crossing.” Econometrica, 78, 1093–1125. [796]
Chernozhukov, V., Fernández-Val, I., and Weidner, M. (2018), “Network and Panel Quantile Effects Via Distribution Regression,” arXiv no. 1803.08154. [783]
Chiang, C. T., Rice, J. A., and Wu, C. O. (2001), “Smoothing Spline Estimation for Varying Coefficient Models With Repeatedly Measured Dependent Variables,” Journal of the American Statistical Association, 96, 605–619. [796]
Dhaene, G., and Jochmans, K. (2015), “Split-Panel Jackknife Estimation of Fixed-Effect Models,” The Review of Economic Studies, 82, 991–1030. [790]
Farcomeni, A. (2012), “Quantile Regression for Longitudinal Data Based on Latent Markov Subject-Specific Parameters,” Statistics and Computing, 22, 141–152. [783]
Fan, J., and Zhang, I. T. (2000), “Functional Linear Models for Longitudinal Data,” Journal of the Royal Statistical Society, Series B, 62, 303–322. [796]
Fan, J., and Zhang, W. (1999), “Statistical Estimation in Varying Coefficient Models,” The Annals of Statistics, 27, 1491–1518. [796]
Fernández-Val, I. (2005), “Bias Correction in Panel Data Models With Individual Specific Parameters,” mimeo, Department of Economics, Boston University, Boston, MA. [789,790]
Frumento, P. (2021), “qrcm: Quantile Regression Coefficients Modeling,” R Package Version 3.0, available at http://CRAN.R-project.org/package=qrcm. [796]
Frumento, P., and Bottai, M. (2016), “Parametric Modeling of Quantile Regression Coefficient Functions,” Biometrics, 72, 74–84, DOI: 10.1111/biom.12410. [783,784,787,788,791,795]
——— (2017), “Parametric Modeling of Quantile Regression Coefficient Functions With Censored and Truncated Data,” Biometrics, 73, 1179–1188, DOI: 10.1111/biom.12675. [783,784,785,791,795]
Geraci, M., and Bottai, M. (2007), “Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution,” Biostatistics, 8, 140–154. [783,796]
——— (2014), “Linear Quantile Mixed Models,” Statistics and Computing, 24, 461–479. [783,796]
Gilchrist, W. (2000), Statistical Modeling With Quantile Functions, Boca Raton: FL: Chapman & Hall. [785]
Hahn, J., and Newey, W. (2004), “Jackknife and Analytical Bias Reduction for Nonlinear Panel Models,” Econometrica, 72, 1295–1319. [789,790]
