Chapter 8
Multi-dimensional Panels in Quantile Regression Models

Antonio F. Galvao and Gabriel V. Montes-Rojas

Abstract This chapter studies estimation and inference methods for multi-dimensional quantile regression panel data models. First, we discuss the fixed effects (FE) model. This model imposes a relatively restrictive asymptotic condition on the growth of the time series dimension relative to the cross section dimension. Nevertheless, extending the FE to three or more dimensions allows for larger data availability, and might help to relax the stringent condition on the time series. We also present a model for the smoothed FE quantile regression case. Second, we present a random effects (RE) model. This model has the advantage of allowing for small time-series dimension. Finally, we present a correlated RE model. In this case, the unobservable individual-specific effects are modeled as a function of observables and a disturbance.

8.1 Introduction

Standard panel data consisting of observations across time for different individuals allow the possibility of controlling for unobserved individual heterogeneity. Such heterogeneity can be an important phenomenon, and failure to control for it may result in misleading inference. This problem is particularly severe when the unobserved heterogeneity is correlated with explanatory variables. Recently, rich panel data sets have become widely available and very popular. They provide a large number of data points, allow analysis of the dynamics of adjustment, as well as control for individual specific heterogeneity.

Quantile regression (QR) models have provided a valuable tool in economics and statistics as a way of capturing heterogeneous effects that covariates may have on the

Antonio F. Galvao
University of Iowa.

Gabriel V. Montes-Rojas
Universitat Autònoma de Barcelona.

outcome of interest, exposing a wide variety of forms of conditional heterogeneity under weak distributional assumptions. Koenker (2004) introduced a general approach to the estimation of QR for panel data models. The panel QR has attracted considerable interest in both the theoretical and applied literatures. It allows one to explore a range of conditional quantiles, thereby exposing a variety of forms of conditional heterogeneity, and to control for unobserved individual-specific effects. Controlling for individual heterogeneity, while exploring heterogeneous covariate effects within the QR framework, offers a more flexible approach to the analysis of panel data than that afforded by the classical fixed and random effects mean-based estimation. QR panel data models are able to capture these two types of heterogeneity in a single framework.

The extension of the two-dimension to the three or higher-dimensional panel data framework has implications for modeling, estimation and inference of conditional quantile models. In turn, these depends on the nature of the multi-dimensional setting, i.e., nested or non-nested, and the type of estimation and inference analyses to be implemented. Moreover, whether the individual-specific effects are correlated or not with the covariates, and at which level, are important elements for the analysis of panel data QR. This chapter studies panel data QR models for multi-dimensional panels and looks into three-dimensional (and higher-dimensional) settings. We concentrate on the standard linear models and discuss the multi-dimensional panels for both fixed and random effects models, and also for correlated random-effects models.

First, we discuss the fixed effects (FE) QR model. The FE-QR allows for individual-specific effects in which no parametric assumptions on the relationship between the specific effects and the covariates are made. Unfortunately, the standard FE-QR estimator is subject to the incidental parameters problem. In addition, there is no general transformation that can suitably eliminate the specific effects in the QR model. Thus, it has been customary to impose a relatively restrictive condition on the growth of the time series dimension relative to the cross section dimension. Nevertheless, extending the standard FE to three dimensions allows for larger data availability, and might help to relax the stringent condition on the time series. In this case, it is even possible that the time dimension is fixed, while the other two dimensions satisfy alternative requirements for asymptotic analysis.

Second, we present the random effects (RE) QR model. The RE-QR model imposes that the specific components are independent of the regressors. In spite of this restriction, in the RE-QR model the unobserved specific effects affect the unobservable variable, which induces heterogeneity across the conditional quantile function of the dependent variable. In addition, the RE-QR has the advantage of allowing for time-invariant regressors, and allows the time-series dimension to be small and fixed.

Finally, we briefly consider the correlated random effects (CRE) QR model. In this case, the unobservable individual-specific effects are modeled as a function of observables and a disturbance. In addition, we will suggest specific guidelines for practitioners in applied work.

The chapter is organized as follows. Section 8.2 describes the fixed effects models considering individual-specific heterogeneity in multi-dimensional panels and analyzes these panel data structures. Section 8.3 studies random-effects models, while Sect. 8.4 explores correlated random-effects frameworks. Finally, Sect. 8.5 summarizes some specific issues for practitioners.

8.2 Fixed Effects Models

In this section we consider a multi-dimensional FE-QR model. In particular, we present a three-dimensional panel data set where the variables of the model are observed along three indices given by the index set (i, j, t) where $i \in \{1, ..., N_1\}$, $j \in \{1, ..., N_2\}$ and $t \in \{1, ..., T\}$, respectively. A FE-QR model with individual-specific and time-specific effects can be written as

$$Q_\tau(y_{ijt}|x_{ijt}, \alpha_i, \gamma_j, \lambda_t) = x_{ijt}\beta(\tau) + \alpha_i(\tau) + \gamma_j(\tau) + \lambda_t(\tau), \tag{8.1}$$

where y_{ijt} is a dependent variable, x_{ijt} is a p-dimensional vector of explanatory variables, α_i and γ_j are the i-th and j-th individual-specific effects, respectively, λ_t the time-specific effect, and $Q_\tau(y_{ijt}|x_{ijt}, \alpha_i, \gamma_j, \lambda_t)$ is the conditional τ-quantile of y_{ijt} given $(x_{ijt}, \alpha_i, \gamma_j, \lambda_t)$. For future reference, we will define $\pi_{ijt} := (\alpha_i, \gamma_j, \lambda_t)$, $\pi_{ij} := (\alpha_i, \gamma_j)$, and $N = N_1 \cdot N_2$. This notation simplifies the discussion on the asymptotic properties, since it implicitly allows us to write $N \to \infty$ when both N_1 and N_2 diverge to infinity, or when one of these dimensions is fixed and the other grows to infinity. In practice, it is often the case that only one individual dimension is large (e.g., firm-employee linked data when the number of employees is much larger than the number of firms). Thus, one dimension, say N_1, might be small or considered fixed, while the other(s), say N_2, is considered large. Therefore, given the specific model of interest to establish the asymptotic properties of the desired estimator, one may consider different scenarios:

$$\begin{aligned}
(i)&\ \ N_1 \to \infty \text{ and } N_2 \text{ fixed},\ T \to \infty,\\
(ii)&\ \ N_1 \text{ fixed and } N_2 \to \infty,\ T \to \infty,\\
(iii)&\ \ N_1 \to \infty \text{ and } N_2 \to \infty,\ T \text{ fixed},\\
(iv)&\ \ N_1 \to \infty \text{ and } N_2 \to \infty,\ T \to \infty.
\end{aligned} \tag{8.2}$$

Model (8.1) can be written as

$$y_{ijt} = x_{ijt}\beta(\tau) + \alpha_i(\tau) + \gamma_j(\tau) + \lambda_t(\tau) + \varepsilon_{ijt}(\tau), \tag{8.3}$$

where $\varepsilon_{ijt}(\tau)$ has a zero conditional τ-quantile given $(x_{ijt}, \alpha_j, \gamma_j, \lambda_t)$. In general, each α_i, γ_j, λ_t and β can depend on τ, but we assume τ to be fixed throughout the section and suppress such dependence for notational simplicity whenever there is no confusion. Model (8.3) assumes that each individual-specific i and j effect enters additively in a linear model.

Koenker (2004) defines the conditional τ-quantile of interest for the dependent variable y conditional on x, for $\tau \in (0,1)$. Two different models can be proposed using this set-up depending on the interpretation of the individual-specific effects. The first is a model in which individual effects do not vary across τ as in Koenker (2004). In this case, the multi-dimensional effects have a pure *location* shift effect on the conditional quantiles. In this case, the pair (i,j) contains intrinsic characteristics which are assumed to be constant when studying conditional heterogeneity. Thus covariate heterogeneity is analyzed at a different level to individual heterogeneity. The second is a model in which individual effects are τ-specific as in Kato et al. (2012). This is a more flexible approach in which (i,j) effects are allowed to vary across the conditional heterogeneity. This is called the *distributional* shift model. However, the QR restrictions on estimation and asymptotic properties reveal that the large T requirement applies to both models, and the choice of each model is a trade-off between flexibility and degrees of freedom (i.e., number of observations with respect to the number of parameters, including the individual effects, to be estimated).

The location conditional τ-quantile of interest for the dependent variable y conditional on x is

$$Q_\tau(y_{ijt}|x_{ijt}) = \beta(\tau)' x_{ijt} + \pi_{ijt}, \tag{8.4}$$

in which π has a pure location shift effect on the conditional quantiles. This quantile model assumes the restrictions $Q_{\tau 1}(y_{ijt}|x_{ijt}) - Q_{\tau_2}(y_{ijt}|x_{ijt}) = (\beta(\tau_1) - \beta(\tau_2))' x_{ijt}$ for all $\tau_1, \tau_2 \in (0,1)$, that is, covariate heterogeneity is present only through changes in the slope parameters β.

The distributional conditional τ-quantile of interest for the dependent variable y conditional on x is

$$Q_\tau(y_{ijt}|x_{ijt}) = \beta(\tau)' x_{ijt} + \pi_{ijt}(\tau), \tag{8.5}$$

in which π has a location-scale shift effect on the conditional quantiles. This quantile model assumes that $Q_{\tau_1}(y_{ijt}|x_{ijt}) - Q_{\tau_2}(y_{ijt}|x_{ijt}) = (\beta(\tau_1) - \beta(\tau_2))' x_{ijt} + \pi_{ijt}(\tau_1) - \pi_{ijt}(\tau_2)$, for all $\tau_1, \tau_2 \in (0,1)$, that is, covariate heterogeneity is present through changes in the slope parameters β and ij-specific intercepts π.

For multi-dimensional panels, we can also consider a "mixed" model in which some intercept parameters vary with τ, while others do not. The choice of the τ-specific and τ-invariant components would depend on the nature of the covariate heterogeneity to be studied.

It is generally assumed that the innovation term ε is independent across individuals, which applied to our case means independence across i and j, but not identically distributed. If the disturbances are assumed to be identically distributed, then $\beta(\tau) = \beta, \forall \tau \in (0,1)$, i.e., the slope parameters are equal across quantiles, all the conditional quantiles are parallel and they only change depending on the location. In general, however, a more flexible model allows for ε to be dependent on the conditioning set $(x, \alpha, \gamma, \lambda)$, in which case $\beta(\tau) \neq \beta$ for some $\tau \in (0,1)$. A canonical example of this situation is the location-scale model in which $\beta(\tau) = \beta + g(x)F_{\varepsilon}^{-1}(\tau)$, for some function $g(\cdot)$ of the covariates.

Now consider the FE-QR model, for which the conditional quantile model of interest is τ-specific,

$$Q_{\tau}(y_{ijt}|x_{it}, \pi_{ij}(\tau)) = \beta(\tau)' x_{ijt} + \pi_{ij}(\tau). \tag{8.6}$$

It is standard in the panel QR literature to treat π as fixed by conditioning on it, as in Hahn and Newey (2004), Fernandez-Val (2005), and Kato et al. (2012). Below we consider the fixed effects estimation of β, which is implemented by treating each individual-specific effect also as a parameter to be estimated. However, given the required estimation of π_{ij}, the FE-QR estimator is, unfortunately, subject to the incidental parameters problem (see Neyman and Scott, 1948; Lancaster, 2000, for a review) and will be inconsistent if the number of individual-specific effects diverges to infinity while the number of time periods T is fixed. It is important to note that, in contrast to mean regression there is no general transformation that can suitably eliminate the specific effects in the QR model. This intrinsic difficulty was recognized by Abrevaya and Dahl (2008), among others, and was clarified by Koenker and Hallock (2000). They remarked that "Quantiles of convolutions of random variables are rather intractable objects, and preliminary differencing strategies familiar from Gaussian models have sometimes unanticipated effects" (p.19).

Therefore, given these difficulties, in the QR panel data literature, it is usual to allow T to increase to infinity at a higher rate than N to achieve consistent estimators. As a result, the standard FE-QR model given by equation (8.6) does not consider a time-specific effect λ_t as a parameter to be estimated. In a multi-dimensional panel, careful consideration of asymptotics of the relative dimensions should be considered. In addition, as we will discuss below, because of the incidental parameters problem, one will be able to control for at most two out of the three specific effects $(\alpha, \gamma, \lambda)$. The allowed specific FE will also depend on the asymptotics considered, as given in cases $(i) - (iv)$ in (8.2) above.

Koenker (2004) and Kato et al. (2012) follow large (N, T) asymptotics (for other recent developments, see, e.g., Galvao, 2011; Galvao et al., 2013; Galvao and Wang, 2015). In the nonlinear and QR literatures, the large panel data asymptotics is used in an attempt to cope with the incidental parameters problem. Canay (2011) proposed a two-step estimator of the common parameters. The difference is that in his model, no individual effect is allowed to change across quantiles, and requires an additional restriction on the conditional average. Graham et al. (2009) show that when $T = 2$ and the explanatory variables are independent of the error term, the FE-QR estimator does not suffer from the incidental parameters problem. However, their argument does not seem to extend to general cases. Rosen (2012) addressed a set identification problem of the common parameters when T is fixed. Chernozhukov et al. (2013) considered identification and estimation of the quantile structural function defined in Imbens and Newey (2009) of a non-separable panel model with discrete explanatory variables. They studied the bounds of the quantile structural function when T is fixed, and the asymptotic behavior of the bounds when T goes to infinity.

In a multi-dimensional setting, the data structure determines the choice of the individual effects that one is able to control for and determines the nature of the model. In several cases, the researcher may be interested in exploring a particular time-invariant covariate set (say across i but not across j). As such, one may choose to explore heterogeneity across a certain dimension and not the other(s). In addition, the FE quantile panel models will produce different models to index heterogeneity depending on the conditional set. For instance, consider the model in equation (8.3). If (α,γ,λ) are controlled for, τ corresponds to an index of heterogeneity in the conditional quantile function of $y|(x,\alpha,\gamma,\lambda)$, which in fact depends on the quantiles of ε. In addition, the choice of the conditional model may depend on the stringent requirements on the time series encountered in the literature for asymptotic analysis. In multi-dimensional panels, additional dimensions are available to the researcher.

8.2.1 Estimation and Implementation

Koenker (2004) and Kato et al. (2012) consider the estimation of the FE using standard QR for a given quantile-τ as follows

$$\left(\hat{\pi},\hat{\beta}\right) = \arg\min_{\pi,\beta} \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} \rho_\tau(y_{ijt} - x'_{ijt}\beta - d'_{ijt}\pi), \tag{8.7}$$

where d_{ijt} is a set of dummy variables that identifies the individual FE for i and j given by π, and $\rho_\tau(u) := u(\tau - I(u<0))$ as in Koenker and Bassett (1978). Note that the coefficients β correspond to the τ-quantile slopes $\beta(\tau)$. The estimation of the regression parameters can be implemented through a QR estimation augmented by the inclusion of the d_{ijt} dummy variables. Note that standard procedures for the estimation of the variance-covariance matrix of this augmented dummy variables estimator are feasible, and then the inference procedures described in the next section could follow from these estimation procedures.

The optimization for solving (8.7) can be very large depending on N_1, N_2 and T. However, as Koenker (2004) observe, in typical applications, the design matrix is very sparse. Standard sparse matrix storage schemes only require the space for the non-zero elements and their indexing locations. This considerably reduces the computational effort and memory requirements. Galvao and Wang (2015) address the computational difficulties and implementation problems without sacrificing the desirable asymptotic properties of the FE-QR strategy. They propose an efficient minimum distance QR estimator, which is very simple to implement in practice. This estimator is defined as the weighted average of the specific QR slope estimators, with weights given by the inverses of the corresponding individual variance-covariance matrices. Moreover, the implementation is not affected by the presence of unbalanced data as the dummy variables strategy works for both balanced and unbalanced panels.

The work in Koenker (2004) also introduced an alternative approach to estimate QR models for panel data with FE that may be subject to shrinkage by ℓ_1 regularization methods. It is well known that the optimal estimator for the random effects Gaussian model involves shrinking the individual effects toward a common value. When there is an intercept in the model, this common value can be taken to be the conditional central tendency of the response at a point determined by the centering of the other covariates. In the QR model, this would be some corresponding conditional quantile of the response. Particularly, when N is large relative to T, shrinkage may be advantageous in controlling the variability introduced by the large number of estimated individual-specific parameters. In this case, the model with shrinkage is

$$\left(\hat{\pi}, \hat{\beta}\right) = \arg\min_{\pi,\beta} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{t=1}^{T} \rho_\tau(y_{ijt} - x'_{ijt}\beta - d'_{ijt}\pi) + \eta\,(\sum_{i=1}^{N_1} |\alpha_i| + \sum_{j=1}^{N_2} |\gamma_j|)\,, \quad (8.8)$$

where $\eta \geq 0$ is a (scalar) penalty or regularization parameter. Note that for $\eta \to 0$ we obtain the FE estimator described above, while as $\eta \to \infty$ we obtain an estimate of the model purged of the FE. In the multi-dimensional case, the penalty is allowed to be different depending on the specific dimensions. For instance we could consider $(\eta_\alpha, \eta_\gamma)$, where the penalty parameter becomes $\eta_\alpha \sum_{i=1}^{N_1} |\alpha_j| + \eta_\gamma \sum_{j=1}^{N_2} |\gamma_j|$. In this case, we could have different degrees of tolerance for shrinking in the different dimensions. Usually for the dimension in which we believe that only a few FE should be non-zero is where $\eta_{(\cdot)}$ should be the largest.

8.2.2 Inference Procedures

As stated above, in standard FE-QR models, the asymptotic analysis for both models requires the time series dimension, T, to increase to infinity to achieve asymptotically unbiased estimators. In a standard two-dimension FE panel, Kato et al. (2012) show that we are required to impose more restrictive conditions on T than that found in the linear panel data FE literature. They show that a sufficient condition to prove asymptotic normality is $N^2(\log N)^3/T \to 0$, which reflects the fact that the rate of the remainder term of the Bahadur representation of the FE-QR estimator is of order $(T/\log N)^{-3/4}$. The slower convergence rate of the remainder term is due to the non-smoothness of the scores. It is important to note that the growth condition on T for establishing $\sqrt{NT}$-consistency of the FE-QR estimator (or other FE estimators in general) is determined so that it "kills" the remainder term. Thus, the rate of the remainder term is essential in the asymptotic analysis of the FE estimation when N and T jointly go to infinity. This restriction requires the cross-sectional dimensions to grow slower than the time-dimension.

In a multi-dimensional setting, asymptotic valid inference will depend on the growth rate of the sample sizes described in equation (8.2). First, note that scenarios

(i) and (ii) in (8.2) require the same conditions on the sample size as stated in Kato et al. (2012). Nevertheless, although these models still require stringent conditions on N relative to T, they allow the researcher to control for π_{ij} because one of these dimensions is finite. However, since scenarios (i) and (ii) require large T, one is not allowed to control for λ_t.

Second, consider the case (iii) in (8.2). This case is also similar to that in Kato et al. (2012) with two dimensions diverging, but the time series dimension is fixed. In this case, one is able to control for only one individual specific effect, i or j, and the requirements on the sample size growth are imposed on the two individual dimensions relative to each other, such that one of the dimensions takes the role of the time series. Note that in this case, since the time series is given, one is also able to control for λ_t.

Finally, consider the case (iv) in (8.2). In this case, one is able to control for the two individual effects, π_{ij}, but note the time effect λ_t, as T diverges to infinity. However, it is important to note that the number of parameters in $\{(\alpha_i,\gamma_j)\}_{i=1,j=1}^{N_1,N_2}$ is N_1+N_2, and this is in general considerably smaller than $N_1 \cdot N_2$. As the conditions discussed are imposed on N $(= N_1 \cdot N_2)$ relative to T, these requirements are more stringent than those that would be required to estimate N_1+N_2 parameters. The main intuition is that although the number of parameters to be estimated grows with the sample size, in the three-dimension panel, the number of parameters to be estimated (N_1+N_2) is smaller than the sample size $(N = N_1 \cdot N_2)$. Another remark on case (iv) is that, since there are three dimensions in the panel data, one is able to exchange the roles of the indices (i, j, t) and estimate one of the individual effects and the time effect, and hence impose the restriction on the remaining dimension to grow fast relative to the other two.

In summary, in cases (i) and (ii) choosing λ_t to be excluded from model (8.6) is based on the idea that in FE models one mainly wishes to control for individual heterogeneity, which in this case is captured by π_{ij}. In case (iii), we could, however, exclude one component in π_{ij}, say α_i, and let its dimension, say N_1, to increase at a higher rate than the other dimensions, say N_2. In this case, one is able to control for λ_t because T is fixed. Finally, in case (iv), one is able to control for two effects only. In this case, one might be able to control for one individual effect and the time effect by exchanging the roles of the indices when considering the relative sample size growth. Therefore, different asymptotic conditions arise depending on different models. In all cases, we need one particular dimension to grow at a faster rate than the number of parameters to be estimated.

In practice, the asymptotic variance of FE-QR estimators depends on the density of the innovation term. For this estimation, different techniques have been suggested and implemented in the literature to produce a consistent estimation of the variance-covariance matrix.

Now we describe the asymptotic normality of $\hat{\beta}$ for case (iv) in (8.2), which can be obtained as in Kato et al. (2012). The other cases are parallel to the standard FE panel data in Kato et al. (2012). Under some regularity conditions, and $N^2(\log N)^3/T \to 0$, we have that

$$\sqrt{NT}(\hat{\beta}-\beta) \xrightarrow{d} N\{0, \tau(1-\tau)\Gamma^{-1}V\Gamma^{-1}\} .$$

Let

$$\tilde{\pi}_{ij} := \mathrm{E}[f_{ij}(0|x_{ij1})x_{ij1}]/f_{ij}(0) \quad \text{and } \Gamma_N := N^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathrm{E}[f_{ij}(0|x_{ij1})x_{ij1}(x'_{ij1}-\tilde{\pi}'_{ij})] ,$$

and where $f_{ij}(u|x)$ is the density of $u_{ijt} = y_{ijt} - \pi_{ij} - x_{jit}\beta$ conditional on x_{ij1} and $f_{ij}(u)$ is the marginal density of u_{ijt}. Let Γ_N be nonsingular for each N, and the limit $\Gamma := \lim_{N\to\infty}\Gamma_N$ exists and is nonsingular; and let the limit

$$V := \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathrm{E}[(x_{ij1}-\pi_{ij})(x_{ij1}-\pi_{ij})']$$

exist and be nonsingular.

Let $\hat{u}_{ijt} = y_{ijt} - \hat{\pi}_{ij} - x'_{jit}\hat{\beta}$. Kato et al. (2012) propose a kernel estimation procedure of the variance of the slope parameters, $V_{\beta(\tau)}$. Let $\mathscr{K} : \mathbb{R} \to \mathbb{R}$ denote a kernel function (probability density function). Let $\{h_N\}$ denote a sequence of positive numbers (bandwidths) such that $h_N \to 0$ as $N \to \infty$ and use the notation $\mathscr{K}_{h_N}(u) = h_N^{-1}\mathscr{K}(u/h_N)$. Assume that the kernel $\mathscr{K}$ is continuous, bounded and of bounded variation on $\mathbb{R}$, and that $h_N \to 0$ and $\log N/(Th_N) \to 0$ as $N \to \infty$.

Define

$$\hat{V}_{1\beta} := \frac{1}{NT}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\mathscr{K}_{h_n}(\hat{u}_{ijt})x_{ijt}(x_{ijt}-\hat{\pi}_{ij})',$$

$$\hat{V}_{0\beta} := \frac{1}{NT}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}(x_{ijt}-\hat{\pi}_{ij})(x_{ijt}-\hat{\pi}_{ij})',$$

where

$$\hat{\pi}_{ij} := \frac{1}{\hat{f}_{ij}T}\sum_{t=1}^{T}\mathscr{K}_{h_N}(\hat{u}_{ijt})x_{ijt}, \quad \hat{f}_{ij} := \frac{1}{T}\sum_{t=1}^{T}\mathscr{K}_{h_N}(\hat{u}_{ijt}).$$

Then, one can consistently estimate the variance-covariance matrix as

$$\hat{V}_{\beta(\tau)} = \tau(1-\tau)\hat{V}_{1\beta}^{-1}\hat{V}_{0\beta}\hat{V}_{1\beta}^{-1}.$$

In practice, as noted above, the variance-covariance matrix can be implemented from standard QR models with the inclusion of individual-specific dummy variables.

8.2.3 Smoothed Quantile Regression Panel Data

A distinctive feature of FE-QR is that its objective function is not differentiable. Nevertheless, the asymptotic analysis depends on the smoothness of objective functions. Kato et al. (2012) formally established the asymptotic properties of the standard FE-QR estimator. However, they required a restrictive condition, such that T grows faster than N^2, to show the asymptotic normality of the estimator, and did not succeed in deriving the bias. The difficulty in handling the standard QR estimator in panel models is partly explained by the fact that the higher order stochastic expansion of the scores is an essential technical tool in the analysis (Hahn and Newey, 2004; Hahn and Kuersteiner, 2004) but such expansion is difficult to implement in the QR case because the Taylor series method is not directly applicable. It is also important to note that the higher order asymptotic behavior of QR estimators is non-standard and rather complicated (Arcones, 1998; Knight, 1998).

An alternative method proposed by Galvao and Kato (2016) is to slightly modify the QR objective function to make it smooth. While this seems an *ad-hoc* change of the objective function, its asymptotic gains are remarkable. The idea of smoothing non-differentiable objective functions goes back to Amemiya (1982) and Horowitz (1992, 1998). Under suitable regularity conditions, the smoothed FE-QR estimator has an order $O(T^{-1})$ bias and hence its limiting normal distribution has a bias in the mean (even) when N and T grow at the same rate. They propose a one-step bias correction estimator based on the analytic form of the asymptotic bias. This is of particular interest in multi-dimensional settings, where the dimension of the individuals is large.

In an attempt to cope with the incidental parameters problem, Galvao and Kato (2016) adopt a different approach and propose a model where N and T grow at the same rate. Instead of the standard QR estimator, the asymptotic properties of the estimator are defined by a minimizer of a smoothed version of the QR objective function.

Smoothing the QR objective function was employed in Horowitz (1998) to study the bootstrap refinement for inference in conditional quantile models. The basic insight of Horowitz (1998) is to smooth over the indicator function $I(y_{ijt} \leq \pi_{ij} + x'_{it}\beta)$ by using a kernel function. To do so, let $K(\cdot)$ be a kernel function and $G(\cdot)$ be the survival function of $K(\cdot)$, i.e.,

$$\int_{-\infty}^{\infty} K(u)du = 1, \; G(u) := \int_{u}^{\infty} K(v)dv.$$

$K(\cdot)$ is not required to be non-negative. Let $\{h_N\}$ be a sequence of positive numbers (bandwidths) such that $h_N \to 0$ as $N \to \infty$ and write $G_{h_N}(\cdot) = G(\cdot/h_N)$. Note that $G_{h_N}(y_{ijt} - \pi_{ij} - \gamma_j - x'_{ijt}\beta)$ is a smoothed counterpart of $I(y_{ijt} \leq \pi_{ij} + x'_{ijt}\beta)$. Then, we consider the estimator

$$(\hat{\pi}, \hat{\beta}) := \arg\min_{(\pi,\beta)} \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} (y_{ijt} - \pi_{ij} - x'_{ijt}\beta)\{\tau - G_{h_N}(y_{ijt} - \pi_{ij} - x'_{ijt}\beta)\}, \quad (8.9)$$

The estimates $\hat{\beta}$ are the FE smoothed quantile regression (FE-SQR) estimator of β.

Galvao and Kato (2016) investigate the asymptotic properties of the FE-SQR estimator defined by (8.9) and provide conditions under which the FE-SQR estimator is consistent and has a limiting normal distribution with a bias in the mean when N and T grow at the same rate. In particular, assuming that $N/T \to \rho$ for some $\rho > 0$, and under some regularity conditions,

$$\sqrt{NT}(\hat{\beta} - \beta) \xrightarrow{d} N(\sqrt{\rho} b, \Gamma^{-1} V \Gamma^{-1}) , \tag{8.10}$$

where

$$b := \Gamma^{-1} \left[\lim_{N\to\infty} \left\{ \frac{1}{N} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} s_{ij} \left(\omega_{ij}^{(1)} \pi_{ij} - \omega_{ij}^{(2)} + \frac{s_{ij} \omega_{ij}^{(3)} \nu_{ij}}{2} \right) \right\} \right] , \tag{8.11}$$

with

$$s_{ij} := 1/f_{ij}(0) , \quad \pi_{ij} := s_{ij} \mathrm{E}[f_{ij}(0|x_{ij1}) x_{ij1}] ,$$

$$\begin{aligned} \nu_{ij} &:= f_{ij}^{(1)}(0) \pi_{ij} - \mathrm{E}[f_{ij}^{(1)}(0|x_{ij1}) x_{ij1}] , \\ \Gamma_N &:= N^{-1} \textstyle\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathrm{E}[f_{ij}(0|x_{ij1}) x_{ij1}(x_{ij1}' - \pi_{ij}')] , \end{aligned}$$

and the limit $\Gamma := \lim_{N\to\infty} \Gamma_N$, and

$$V := \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} V_{ij}$$

with V_{ij} denoting the covariance matrix of the term

$$T^{-1/2} \sum_{t=1}^{T} \{\tau - I(u_{ijt} \le 0)\}(x_{ijt} - \pi_{ij}) .$$

Moreover,

$$\begin{aligned} \omega_{ij}^{(1)} &:= \sum_{1\le|k|\le T-1} \left(1 - \frac{|k|}{T}\right) \left\{ \tau f_{ij}(0) - \int_{-\infty}^{0} f_{ij,k}(0,u) du \right\} , \\ \omega_{ij}^{(2)} &:= \sum_{1\le|k|\le T-1} \left(1 - \frac{|k|}{T}\right) \Big\{ \tau \mathrm{E}[f_{ij}(0|x_{ij1}) x_{ij1}] \\ &\qquad - \mathrm{E}\left[x_{ij1} \int_{-\infty}^{0} f_{ij,k}(0,u|x_{ij1}, x_{ij,1+k}) du \right] \Big\} , \\ \omega_{ij}^{(3)} &:= \sum_{|k|\le T-1} \left(1 - \frac{|k|}{T}\right) \mathrm{Cov}\{I(u_{ij1} \le 0), I(u_{ij,1+k} \le 0)\} . \end{aligned}$$

The exact form of the term V_{ij} is given by

$$V_{ij} = \sum_{|k|\leq T-1} \left(1 - \frac{|k|}{T}\right) \mathrm{E}\left[\{\tau - I(u_{ij1} \leq 0)\}\{\tau - I(u_{ij,1+k} \leq 0)\} \times \right.$$
$$\left. \times\, (x_{ij1} - \pi_{ij})(x_{ij,1+k} - \pi_{ij})'\right].$$

If there is no time series dependence, i.e., for each i, j, the process $\{(y_{ijt}, x_{ijt}), t = 0, \pm 1, \pm 2, \dots\}$ is i.i.d., then

$$V_{ij} = \tau(1-\tau)\mathrm{E}[(x_{ij1} - \pi_{ij})(x_{ij1} - \pi_{ij})'], \quad \omega_{ij}^{(1)} = 0, \quad \omega_{ij}^{(2)} = 0, \quad \text{and}$$
$$\omega_{ij}^{(3)} = \tau(1-\tau)$$

8.2.3.1 Bias Correction – Analytical Method

As stated in the literature, the problem of the limiting distribution of $\sqrt{NT}(\hat{\beta} - \beta)$ not being centered at zero is that usual confidence intervals based on the asymptotic approximation will be incorrect. In particular, even if b is small, the asymptotic bias can be of moderate size when the ratio N/T is large. In this subsection, we shall consider the bias correction to the FE-SQR estimator.

Consider a one-step bias correction based on the analytic form of the asymptotic bias. Put $\hat{u}_{ijt} := y_{it} - \hat{\pi}_{ij} - x'_{ijt}\hat{\beta}$. The terms $f_{ij} := f_{ij}(0), s_{ij}, \pi_{ij}, \nu_{ij}$ and Γ can be estimated by

$$\hat{f}_{ij} := \frac{1}{T}\sum_{t=1}^{T} K_{h_N}(\hat{u}_{ijt}), \quad \hat{s}_{ij} := \frac{1}{\hat{f}_{ij}}, \quad \hat{\pi}_{ij} := \frac{\hat{s}_{ij}}{T}\sum_{t=1}^{T} K_{h_N}(\hat{u}_{ijt}) x_{ijt},$$
$$\hat{\nu}_{ij} := \frac{1}{T h_N^2}\sum_{t=1}^{T} K^{(1)}(\hat{u}_{ijt}/h_N)(x_{ijt} - \hat{\pi}_{ij}),$$
$$\hat{\Gamma}_N := \frac{1}{NT}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} K_{h_N}(\hat{u}_{ijt}) x_{ijt}(x'_{ijt} - \hat{\pi}'_{ij}),$$

where $K^{(1)}(u) = dK(u)/du$. The estimation of the terms $\omega_{ij}^{(1)}, \omega_{ij}^{(2)}$ and $\omega_{ij}^{(3)}$ is a more delicate issue, since it reduces to the estimation of long run covariances. As in Hahn and Kuersteiner (2004), we make use of a truncation strategy. Define

$$\phi_{ij}(k) := \int_{-\infty}^{0} f_{ij,k}(0,u)du,$$
$$\varphi_{ij}(k) := \mathrm{E}\left[x_{ij1}\int_{-\infty}^{0} f_{ij,k}(0,u|x_{ij1}, x_{ij,1+k})du\right],$$
$$\rho_{ij}(k) := \mathrm{E}[I(u_{ij1} \leq 0) I(u_{ij,1+k} \leq 0)].$$

Since $\phi_{ij}(k) \approx \mathrm{E}[K_{h_N}(u_{ij1}) I(u_{ij,1+k} \leq 0)]$, it can be estimated by

$$\hat{\phi}_{ij}(k) := \frac{1}{T} \sum_{t=\max\{1,-k+1\}}^{\min\{T,T-k\}} K_{h_N}(\hat{u}_{ijt}) I(\hat{u}_{ij,t+k} \leq 0).$$

Similarly, $\varphi_{ij}(k)$ can be estimated by

$$\hat{\varphi}_{ij}(k) := \frac{1}{T} \sum_{t=\max\{1,-k+1\}}^{\min\{T,T-k\}} K_{h_N}(\hat{u}_{ijt}) I(\hat{u}_{ij,t+k} \leq 0) x_{ijt}.$$

The term $\rho_{ij}(k)$ can be estimated by its sample analogue:

$$\hat{\rho}_{ij}(k) := \frac{1}{T} \sum_{t=\max\{1,-k+1\}}^{\min\{T,T-k\}} I(\hat{u}_{ijt} \leq 0) I(\hat{u}_{ij,t+k} \leq 0).$$

Take a sequence m_N such that $m_N \to \infty$ sufficiently slowly. Then, $\omega_{ij}^{(1)}, \omega_{ij}^{(2)}$ and $\omega_{ij}^{(3)}$ can be estimated by

$$\begin{aligned}
\hat{\omega}_{ij}^{(1)} &:= \sum_{1 \leq |k| \leq m_N} \left(1 - \frac{|k|}{T}\right) \{\tau \hat{f}_{ij} - \hat{\phi}_{ij}(k)\}, \\
\hat{\omega}_{ij}^{(2)} &:= \sum_{1 \leq |k| \leq m_N} \left(1 - \frac{|k|}{T}\right) \{\tau \hat{f}_{ij} \hat{\pi}_{ij} - \hat{\varphi}_{ij}(k)\}, \\
\hat{\omega}_{ij}^{(3)} &:= \tau(1-\tau) + \sum_{1 \leq |k| \leq m_N} \left(1 - \frac{|k|}{T}\right) \{-\tau^2 + \hat{\rho}_{ij}(k)\}.
\end{aligned}$$

The bias term b is thus estimated by

$$\hat{b} := \hat{\Gamma}_N^{-1} \left\{ \frac{1}{N} \sum_{i=1}^{N_1} \sum_{i=1}^{N_2} \hat{s}_{ij} \left(\hat{\omega}_{ij}^{(1)} \hat{\pi}_{ij} - \hat{\omega}_{Nij}^{(2)} + \frac{\hat{s}_{ij} \hat{\omega}_{ij}^{(3)} \hat{\nu}_{ij}}{2} \right) \right\}.$$

We define the one-step bias corrected estimator by $\hat{\beta}^1 := \hat{\beta} - \hat{b}/T$. In practice, there is no need to compute the terms $\tau \hat{f}_{ij}$ and $\tau \hat{f}_{ij} \hat{\pi}_{ij}$ in $\hat{\omega}_{ij}^{(1)}$ and $\hat{\omega}_{ij}^{(2)}$, respectively, as they are canceled out by the difference $\hat{\omega}_{ij}^{(1)} \hat{\pi}_{ij} - \hat{\omega}_{ij}^{(2)}$. Additionally, there is no need to use the same kernel and the same bandwidth to estimate β and b.

Galvao and Kato (2016) show that the bias corrected estimator, $\hat{\beta}^1$, has the limiting normal distribution with mean zero and the same covariance matrix as $\hat{\beta}$ as

$$\sqrt{NT}(\hat{\beta}^1 - \beta) \xrightarrow{d} N(0, \Gamma^{-1} V \Gamma^{-1}),$$

when $m_N \to \infty$ such that $m_N^2 (\log N)/(T h_N^2) \to 0$.

8.2.3.2 Jackknife

Galvao and Kato (2016) also consider the half-panel jackknife method originally proposed by Dhaene and Jochmans (2015), as an estimator. This method is an automatic way of removing the bias of $\hat{\beta}$. Suppose for a moment that T is even. Partition $\{1,\ldots,T\}$ into two subsets, $S_1 := \{1,\ldots,T/2\}$ and $S_2 := \{T/2+1,\ldots,T\}$. Let $\hat{\beta}_{S_l}$ be the FE-SQR estimate based on the data $\{(y_{ijt},x_{ijt}), 1 \leq i \leq N_1, 1 \leq j \leq N_2, t \in S_l\}$ for $l = 1,2$. The half-panel jackknife estimator is defined as $\hat{\beta}_{1/2} := 2\hat{\beta} - \bar{\beta}_{1/2}$, where $\bar{\beta}_{1/2} := (\hat{\beta}_{S_1} + \hat{\beta}_{S_2})/2$. For simplicity, suppose for a moment that we use the same bandwidth to construct $\hat{\beta}$ and $\hat{\beta}_{S_l}$ $(l = 1,2)$. Then, from the asymptotic representation of the FE-SQR estimator, it can be shown that under some regularity conditions

$$\sqrt{NT}(\hat{\beta}_{1/2} - \beta) \xrightarrow{d} N(0, \Gamma^{-1}V\Gamma^{-1}).$$

The half-panel does not require the non-parametric estimation of the bias term and at the same time is easy to implement empirically.

8.3 Random Effects Models

Random effects (RE) models have recently been considered in the QR panel data framework. As noted in Chap. 2, RE models have two main advantages. First, the RE approach does not suffer from the incidental parameters problem, that is the number of parameters to take into account does not increase with the sample size (i.e., N). This is an important restriction to be lifted for QR models as noted previously, because for FE-QR models the existing sufficient conditions under which the asymptotic bias of the FE-QR vanishes require $T >> N$. For the general multi-dimensional setting, this is an important restriction for models in which the number of intercepts is of the order $O(N_1 + N_2)$ or $O(N_1 \cdot N_2)$. Koenker (2004) argues that the latter "would certainly be useful for groups of individuals: a distributional shift for men versus women, or for blacks versus whites. However, in most applications the $[T]$, the number of observations in the time series, would be relatively modest and then it is quite unrealistic to attempt to estimate a τ-dependent, distributional, individual effect" (p.76). In most applications, the time series dimension T is indeed relatively small compared to the number of individuals.

Second, the RE model also makes possible the identification of parameters associated with individual (and time) invariant variables. In a multi-dimensional framework, this may be of interest for the applied researcher that wants to control for (i,j,t) heterogeneity while exploring covariate heterogeneity across conditional quantiles.

Galvao and Poirier (2015) develop a RE model for QR panel data with time invariant regressors. They establish identification, and develop practical estimation and inference procedures. In this section, we extend the RE model to the multi-dimensional context and apply a simple pooled QR estimator to estimate the coefficients of interest and establish its statistical properties. We also suggest a cluster robust variance-covariance matrix estimator for inference, and establish its uniform consistency. The RE model is interesting because it allows the researcher to control for time-invariant regressors, as well as use small a panel where the time dimension, T, is small.

8.3.1 Model

Consider now a linear RE-QR model with scalar multi-dimensional specific effects (for simplicity, we follow the notation of Chap. 1 rather than Chap. 2). Let z_{ij} be a set of covariates that does not vary across t.

Following Galvao and Poirier (2015), we begin the discussion with a random coefficients representation of the form

$$y_{ijt} = c(U_{ijt}) + x_{ijt}'\beta(U_{ijt}) + z_{ij}'\delta(U_{ijt}), \tag{8.12}$$

where U_{ijt} represents the heterogeneity in responses and can depend on both ε_{ijt} and π_{ij}, as

$$U_{ijt} \equiv U(\pi_{ij}, \varepsilon_{ijt}), \tag{8.13}$$

with $U(\cdot,\cdot)$ being a scalar and unspecified non-parametric function. Note that equation (8.13) allows the unobserved heterogeneity to depend on both the independent unobserved component, ε_{ijt}, and the individual-specific components, π_{ij}, in an unrestricted form. The functions $c(\cdot), \beta(\cdot)$ and $\delta(\cdot)$ in (8.12) quantify the distributional effects for the intercept, and the time-varying and time-invariant regressors, x_{ijt} and z_{ij} respectively. Note that for the RE-QR models, since the time series dimension T is fixed, one can easily include a time-specific effect λ_t as regressors with corresponding parameters to be estimated inside the vector x_{ijt} in equation (8.12), hence we only consider π_{ij} in equation (8.13).

The RE assumption in standard linear mean panel data models restricts the unobserved component, π_{ij}, to being uncorrelated with all regressors. We generalize this assumption to the model in (8.12)–(8.13) by assuming the following independence condition

$$(\pi_{ij}, \varepsilon_{ijt}) \perp\!\!\!\perp (x_{ijt}, z_{ij}). \tag{8.14}$$

The stronger independence assumption in (8.14) is used due to the non-linearity in π_{ij} of equations (8.12)–(8.13). Thus the unobserved heterogeneity $U(\pi_{ij}, \varepsilon_{ijt})$ is independent of (x_{ijt}, z_{ij}), which gives rise to the following quantile representation

$$Q_\tau(y_{ijt}|x_{ijt}, z_{ij}) = c(\tau) + x_{ijt}'\beta(\tau) + z_{ij}'\delta(\tau)\ , \tag{8.15}$$

where the presence of τ on the right-hand side follows from our previous normalization of U_{ijt} and from $Q_\tau(U_{ijt}|x_{ijt}, z_{ij}) = Q_\tau(U_{ijt}) = \tau$. Equation (8.15) establishes the linear RE-QR model, given equations (8.12)–(8.13), and condition (8.14). For notation convenience define $w_{ijt} = [1, x'_{ijt}, z'_{ij}]'$ and $\theta(.) = [c(.), \beta(.), \delta(.)]$.

Before we present the estimation, it is important to discuss the differences between the FE-QR and the RE-QR. Galvao and Poirier (2015) show that the relationship between RE and FE is more delicate for quantile models than for standard mean-regression models. In a linear panel model, traditional conditional mean FE and RE estimation are based on the same linear model, which often takes the form

$$y_{ijt} = x'_{ijt}\beta + \pi_{ij} + \varepsilon_{ijt}.$$

Under FE, i.e., $\text{Cov}(\alpha_{ij}, x_{ijt}) \neq 0$, the "within" estimator can recover β. On the other hand, if the RE assumption of $\text{Cov}(\alpha_{ij}, x_{ijt}) = 0$ is assumed, the pooled regression estimator will also be consistent for β. However, in QR models, the FE and RE models differ substantially.

To see this difference, consider the additive-in-π_{ij} linear FE-QR model, which can be represented as

$$y_{it} = \pi_{ij} + x'_{ijt}\beta(\varepsilon_{ijt}). \tag{8.16}$$

There are two important points regarding model (8.16). First, note that under FE or RE, the *conditional* model yields $Q_\tau(y_{ijt}|x_{ijt}, \pi_{ij}) = \pi_{ij} + x'_{ijt}\beta(\tau)$, a linear expression. Under some regularity conditions, and no restriction on the relationship between π_{ij} and x_{ijt}, Kato et al. (2012) show that $\beta(\tau)$ can be estimated consistently by a QR with individual-specific dummy variables when T is large. Note that the inclusion of individual-specific dummy variables precludes one from having time-invariant regressors. Nevertheless, even if $\pi_{ij} \perp\!\!\!\perp x_{ijt}$ holds in (8.16) a simple RE estimator would be unable to consistently estimate $\beta(\tau)$ since the conditional quantile function would be misspecified if we do not condition on the individual specific effects π_{ij}. This implies that the pooled QR and the FE-QR estimators estimate different quantities.

Second, note that there are differences between $Q_\tau(y_{ijt}|x_{ijt}, \pi_{ij})$, the *conditional* model, and $Q_\tau(y_{ijt}|x_{ijt})$, the *marginal* model. Even if $\pi_{ij} \perp\!\!\!\perp x_{ijt}$ holds in (8.16), the conditional quantile of y_{ijt} given x_{ijt} might not be linear in x_{ijt} since $Q_\tau(y_{ijt}|x_{ijt}) \neq Q_\tau(\pi_{ij}) + x'_{ijt}\beta(\tau)$, because the quantile of a sum is generally different from the sum of the quantiles by the non-linearity of the quantile operator.[1] Again, the pooled QR and the FE-QR estimators estimate different quantities.

In equations (8.12)–(8.13), consider an alternative model which is non-additive in the individual-specific effect, α_i, as

$$y_{ijt} = x'_{ijt}\beta(U(\pi_{ij}, \varepsilon_{ijt})).$$

Under assumption (8.14), the conditional quantile of y_{ijt} given x_{ijt} in the above equation, i.e., the marginal model, is linear in x_{ijt} and $\beta(\tau)$. Again, the conditional

[1] The conditional quantile of the sum will be equal to the sum of conditional quantiles if conditional co-monotonicity between $\pi_{ij}i$ and $x'_{it}\beta(\varepsilon_{ijt})$ holds conditional on x_{ij}. This is ruled out by the conditional independence of π_{ij} and ε_{ijt} given x_{ij} since co-monotonic variables cannot be independent (see Galvao and Poirier (2015) for a proof of this result).

model will yield a different effect of x_{ijt} on y_{ijt}, so a FE estimator with individual-specific dummy variables would not recover the same coefficients on x_{ijt} due to the non-additivity of π_{ij}. Thus, the FE-QR and RE-QR estimators will converge to different quantities, since they must rely on different modeling assumptions, which stands in contrast to the linear mean-regression panel case, where both the FE and RE estimators are consistent. This is due to the fact that under RE, the marginal and conditional models for linear, mean-regression panels yield the same effect of x_{ijt} on y_{ijt},[2] while these effects differ for all models considered here.

Another feature of the non-additive RE-QR model is that a failure of the RE assumption (8.14) will imply that the conditional quantile of y_{ijt} is no longer linear in w_{ijt}, because the composite unobserved heterogeneity term U_{ijt} will generally be correlated with w_{ijt}. We can then write the conditional quantile of y_{ijt} as follows

$$\begin{aligned} Q_\tau(y_{ijt}|w_{ijt}) &= w'_{ijt}\theta(Q_\tau(U_{ijt}|w_{ijt})) \\ &\equiv w'_{ijt}\tilde{\theta}(\tau;w_{ijt}), \end{aligned}$$

where $\tilde{\theta}(\tau;w_{ijt})$ is a non-parametric function of τ and w_{ijt}. Graham et al. (2015) discuss the non-parametric identification and estimation of this model, and more specifically of the unconditional quantile effect $\theta(\tau)$.[3]

8.3.2 Estimation and Implementation

Based on the identification condition (8.14) and the model given in (8.15), a simple pooled QR estimator for $\theta(\tau) = [c(\tau), \beta(\tau)', \delta(\tau)']'$ can be employed. The estimator is defined as follows:

$$\hat{\theta}(\tau) \equiv \arg\min_\theta \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} \rho_\tau(y_{ijt} - w'_{ijt}\theta), \tag{8.17}$$

where $\rho_\tau(u) \equiv \{\tau - 1(u \leq 0)\}u$ is, again, the check function (Koenker and Bassett, 1978). Therefore, the practical estimation procedure for the coefficients of interest is very simple and can thus be implemented through standard QR estimation. First, one stacks the data, and second, applies a simple QR. Nevertheless, given that the individual effects induce clustering, the inference needs to be adjusted. We describe inference procedures in the next section.

---

[2] This can be seen from $\mathrm{E}[y_{ijt}|x_{ijt}, \pi_{ij}] = \pi_{ij} + x'_{ijt}\beta$ and $\mathrm{E}[y_{ijt}|x_{ijt}] = \mathrm{E}[\pi_{ij}] + x'_{ijt}\beta$.

[3] These differences between the RE and FE models in the QR case make testing for the presence of RE very important in the QR context. Galvao and Poirier (2015) provide such a test.

8.3.3 Inference Procedures

Under some standard regularity conditions, Galvao and Poirier (2015) derive the asymptotic normality of the RE-QR estimator as follows. For a given quantile τ of interest, as $N \to \infty$, $\hat{\theta}(\cdot)$

$$\sqrt{N}(\hat{\theta}(\tau) - \theta(\tau)) \xrightarrow{d} N(0, \Gamma(\tau)^{-1} V(\tau) \Gamma(\tau)^{-1}),$$

where $\Gamma(\tau) \equiv \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} w_{ijt} w_{ijt}' f_{y_{ijt}}(w_{ijt}'\theta(\tau)|w_{ij})\right]$, and

$$V(\tau, \tau') = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} \mathrm{E}\left[(1(v_{it}(\tau) \leq 0) - \tau)(1(v_{is}(\tau') \leq 0) - \tau')X_{it}X_{is}'\right]. \quad (8.18)$$

The existence of the RE in the model generates cluster-dependence, and thus the standard errors require a cluster-robust variance-covariance matrix estimation.

For given quantiles of interest, the variance-covariance matrix of $\hat{\theta}(\tau)$ is

$$\Gamma(\tau)^{-1} V(\tau) \Gamma(\tau)^{-1}$$

with components $\Gamma(\tau) = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} w_{ijt} w_{ijt}' f_{y_{ijt}}(w_{ijt}'\theta(\tau)|w_{ij})\right]$, and (8.18) can be rewritten as

$$\begin{aligned} V(\tau) &= \mathrm{E}\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}(1(v_{ijt}(\tau) \leq 0) - \tau)(1(v_{ijs}(\tau) \leq 0) - \tau) w_{ijs} x_{ijt}'\right] \\ &= \frac{\tau(1-\tau)}{T^2}\sum_{t=1}^{T} \mathrm{E}[w_{ijt} w_{ijt}'] \\ &\quad + \frac{1}{T^2}\sum_{s \neq t} \mathrm{E}\left[\mathrm{Cov}(1(v_{ijs}(\tau) \leq 0), 1(v_{ijt}(\tau) \leq 0)|w_{ijs}, w_{ijt}) w_{ijs} w_{ijt}'\right], \end{aligned}$$

where $v_{it}(\tau) \equiv y_{ijt} - (c(\tau) + x_{ijt}'\beta(\tau) + z_{ij}'\gamma(\tau))$ and $f(.)$ is the conditional density of y_{ijt} given $w_{ij} = [w_{ij1}, w_{ij2}, ..., w_{ijT}]$. Note that in the second component, the second term disappears if there is no intra-unit dependence of the QR residuals. Thus, using simple standard errors for the pooled QR estimator without correcting for the cluster-dependence will produce incorrect inference unless the second term is zero.

To conduct practical inference, consider

$$\hat{\Gamma}(\tau) = \frac{1}{NT}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} w_{ijt} w_{ijt}' \frac{1}{h_N} K\left(\frac{\hat{v}_{ijt}(\tau)}{h_N}\right),$$

where $\hat{v}_{ijt}(\tau)$ are the estimated residuals, and $K(\cdot)$ is a kernel function of bounded variation, and h_N is a bandwidth. This is a variant of the Powell (1986) kernel estimator for QR in cross-sectional models. The component $V(\tau)$ can be estimated by

$$\hat{V}(\tau) = \frac{\tau(1-\tau)}{NT^2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{t=1}^{T} w_{ijt} w'_{ijt}$$
$$+ \frac{1}{NT^2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{s \neq t} w_{ijs} w'_{ijt} (1(\hat{v}_{ijs}(\tau) \leq 0) - \tau)(1(\hat{v}_{ijt}(\tau) \leq 0) - \tau).$$

8.4 Correlated Random Effects Models

Another alternative class of models for QR panel data is the correlated random effects. The correlated RE model of Chamberlain (1982, 1984) views the unobservable individual specific component π_{ij} as a linear projection onto the observables plus a disturbance. The intuition behind these models is that a rich set of covariates is able to explain unobserved heterogeneity and what is left is idiosyncratic noise. This idea has also been implemented in QR panel data models.

Abrevaya and Dahl (2008) introduced an alternative approach to the FE-QR, which estimates QR models for panel data employing the correlated random effects (CRE) model of Chamberlain (1982). The unobservable individual specific effect is modeled as a linear projection onto the observables and a disturbance. Geraci and Bottai (2007) consider a RE approach for a single quantile assuming that the outcome variable is distributed as an asymmetric Laplace distribution conditional on covariates and individual effects. Arellano and Bonhomme (2016) introduce a class of QR estimators for short panels, where the conditional quantile response function of the unobserved heterogeneity is specified as a function of observables. They develop a model general model for nonlinear panel data that covers static and dynamic autoregressive models, models with general predetermined regressors, and models with multiple individual effects. However, the correlated RE requires a specification of the individuals specific effects as a known function of the observables.

Extensions to the multi-dimensional case allow for different covariates to be used for each specific components, and as such, to be able to produce a more accurate model of unobserved effects. Let z_i, w_j and b_j be covariate sets that only vary on a given dimension, i, j, and t, possibly nested within x_{ijt}, and let (a_i, g_j, r_t) be unobserved component such that

$$\alpha_i = \lambda_1' z_i + a_i,$$
$$\gamma_j = \lambda_2' w_j + g_j,$$
$$\lambda_t = \lambda_3' b_t + r_t.$$

Following Abrevaya and Dahl (2008), the strategy is to replace them into equation (8.5) to obtain an explicit model of the quantiles as

$$Q_\tau(y_{ijt}|x_{ijt}) = \beta(\tau)' x_{ijt} + \lambda_1(\tau)' z_i + \lambda_2(\tau)' w_j + \lambda_3' b_t.$$

Arellano and Bonhomme (2016) develop an estimation strategy for general nonlinear panel models. They specify outcomes y_{ijt} as a function of covariates x_{ijt} and latent heterogeneity π_{ij} as

$$y_{ijt} = \sum_{k=1}^{K_1} \theta_k(U_{ijt}) g_k(x_{ijt}, \pi_{ij}),$$

and similarly specify the dependence of π_{ij} on covariates

$$\pi_{ij} = \sum_{k=1}^{K_2} \delta_k(V_{ij}) h_k(x_{ijt}),$$

where $U_{ij1},...,U_{ijT},V_{ij}$ are independent uniform random variables, and g() and h() belong to some family of functions. Outcomes and heterogeneity are monotone in U_{ijt} and V_{ij}, respectively, so the above models correspond to conditional quantile functions. This is a correlated RE model that can become arbitrarily flexible as K_2 increases. For the multi-dimensional case, this could be made as

$$\alpha_i = \sum_{k=1}^{K_2} \delta_h^{\alpha}(V_{\alpha i}) h_k^{\alpha}(x_{ijt}),$$

$$\gamma_j = \sum_{k=1}^{K_3} \delta_h^{\gamma}(V_{\gamma j}) h_k^{\gamma}(x_{ijt}),$$

where heterogeneity is modeled in a different way for each dimension.

8.5 Specific Guidelines for Practitioners

QR has attracted considerable interest in econometrics and statistics. It offers an easy-to-implement method to estimate conditional quantiles. Recently, there has been a growing literature on estimation and testing using QR for panel data models. Panel QR has provided a valuable method of statistical analysis of the heterogeneous effects of policy variables.

Nevertheless, as discussed above, one particular difficulty in QR panel data models, both for fixed and random effects models, is that the asymptotic variance of QR estimators depends on the density of the innovation term, and it is not easy to compute in practice. We have presented several procedures for estimating the variance-covariance matrix in their corresponding models, all of them with a kernel implementation whenever the density is involved. This in turn depends on the specific multi-dimensional setting. Additional research is needed to evaluate the relative performance of each procedure. By selecting a specific model depending on the dimension to be considered as fixed or random, we are in effect modeling different quantiles, that is, different models to analyze the heterogeneity of the effects of covariates on an outcome variable. This should be guided by the specific interest of the empirical analysis, in terms of why we are studying quantile heterogeneity.

Inference procedures and confidence interval construction can be greatly simplified by using bootstrap methods. Specific designs for different QR problems may be guided by the bootstrap results developed in the mean regression case. In particular, different bootstrapping procedures for panel data models, as in Kapetanios (2008), can be easily adapted to the multi-dimensional setting. Galvao and Montes-Rojas (2015) argue that bootstrapping techniques greatly simplify the variance-covariance estimation.[4] They propose to construct confidence intervals for the parameters of interest using percentile bootstrap with pairwise resampling. In practice, FE and RE QR parameters' point estimates can thus be implemented using standard QR codes available in econometric softwares, that work for both balanced and unbalanced panel data, and different bootstrapping techniques could be adapted for either FE or RE.

As discussed above, panel data QR estimators' consistency and other asymptotic properties rely on the dimension of the heterogeneity being described, either as FE or RE. In the multi-dimensional setting, this may exponentially grow depending on the researcher's choice. If for mean-based models this is a serious issue for efficiency reasons (i.e., degrees of freedom), the asymptotic results above should suggest caution in QR models with large dimensions. In particular, smoothed and/or mixed models should be considered to reduce potential asymptotic bias. As a practical example, if the researcher can choose which dimension is potentially correlated with covariates of interest and which one is not, then the former could be considered as a fixed-parameter to be estimated, with the corresponding incidental parameter problem, and the other could be modeled as a random effect. Galvao and Poirier (2015) test for RE vs. FE models could help in this direction.

References

Abrevaya, J. and Dahl, C. M. (2008). The effects of birth inputs on birthweight: Evidence from quantile estimation on panel data. *Journal of Business and Economic Statistics*, 26:379–397.

[4] Bootstrapping techniques have ~~extensively been~~ used to construct confidence intervals for QR in the cross-sectional context. Buchinsky (1995) uses Monte Carlo simulation to study several estimation procedures of the asymptotic covariance matrix in quantile regression models, and the results favor the bootstrap design. Hahn (1995) shows that the construction of confidence intervals based on the QR estimators can be greatly simplified by using bootstrapping. Moreover, the confidence intervals constructed by the bootstrap percentile method have asymptotically correct coverage probabilities. Horowitz (1998) proposes bootstrap methods for median regression models. Feng et al. (2011) propose an adaptation wild bootstrap methods for QR. Wang and He (2007) develop inference procedures based on rank-score tests with RE. In the panel data FE context, Abrevaya and Dahl (2008) use bootstrapping for constructing confidence intervals in the QR panel data.

Amemiya, T. (1982). Two stage least absolute deviations estimators. *Econometrica*, 50:689–711.

Arcones, M. A. (1998). Second order representations of the least absolute deviation regression estimator. *Annals of the Institute of Statistical Mathematics*, 50:87–117.

Arellano, M. and Bonhomme, S. (2016). Nonlinear panel data estimation via quantile regressions. *Econometrics Journal*, forthcoming.

Buchinsky, M. (1995). Estimating the asymptotic covariance matrix for quantile regression models: A Monte Carlo study. *Journal of Econometrics*, 68:303–338.

Canay, I. A. (2011). A simple approach to quantile regression for panel data. *Econometrics Journal*, 14:368–386.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18:5–46.

Chamberlain, G. (1984). Panel data. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, pages 1248–1313. Amsterdam: North-Holland.

Chernozhukov, V., Fernandez-Val, I., Hahn, J., and Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81:535–580.

Dhaene, G. and Jochmans, K. (2015). Split-panel jacknife estimation of fixed-effect models. *Review of Economic Studies*. forthcoming.

Feng, X., He, X., and Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*, 98(4):995–999.

Fernandez-Val, I. (2005). Bias correction in panel data models with individual specific parameters. Mimeo.

Galvao, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, 164:142–157.

Galvao, A. F. and Kato, K. (2016). Smoothed quantile regression for panel data. *Journal of Econometrics*, 193:92–112.

Galvao, A. F., Lamarche, C., and Lima, L. (2013). Estimation of censored quantile regression for panel data with fixed effects. *Journal of the American Statistical Association*, 108:1075–1089.

Galvao, A. F. and Montes-Rojas, G. V. (2015). On bootstrap inference for quantile regression panel data: A Monte Carlo study. *Econometrics*, 3:654–666.

Galvao, A. F. and Poirier, A. (2015). Quantile regression random effects. Mimeo.

Galvao, A. F. and Wang, L. (2015). Efficient minimum distance estimator for quantile regression fixed effects panel data. *Journal of Multivariate Analysis*, 133:1–26.

Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1):140–154.

Graham, B. S., Hahn, J., Poirier, A., and Powell, J. L. (2015). A quantile correlated random coefficients panel data model. Mimeo.

Graham, B. S., Hahn, J., and Powell, J. L. (2009). The incidental parameter problem in a non-differentiable panel data model. *Economics Letters*, 105:181–182.

Hahn, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory*, 11:105–121.

Hahn, J. and Kuersteiner, G. M. (2004). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. Mimeo.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319.

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica*, 60:505–531.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica*, 66:1327–1351.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equation models without additivity. *Econometrica*, 77:1481–1512.

Kapetanios, G. (2008). A bootstrap procedure for panel datasets with many cross-sectional units. *Econometrics Journal*, 11:377–395.

Kato, K., Galvao, A. F., and Montes-Rojas, G. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170:76–91.

Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *Annals of Statistics*, 26:755–770.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91:74–89.

Koenker, R. and Bassett, G. W. (1978). Regression quantiles. *Econometrica*, 46:33–49.

Koenker, R. and Hallock, K. (2000). Quantile regression: An introduction. Manuscript, University of Illinois at Urbana-Champaign.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95:391–413.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, 32:143–155.

Rosen, A. (2012). Set identification via quantile restrictions in short panels. *Journal of Econometrics*, 166(1):127–167.

Wang, H. and He, X. (2007). Detecting differential expressions in GeneChip microarray studies. *Journal of the American Statistical Association*, 102:104–112.