Random Effects Quantile Regression*

Manuel Arellano — CEMFI, Madrid
Stéphane Bonhomme — CEMFI, Madrid

This draft: 26 February 2013
PRELIMINARY AND INCOMPLETE

Abstract

We introduce a class of linear quantile regression estimators for panel data. Our framework contains dynamic autoregressive models, models with general predetermined regressors, and models with multiple individual effects as special cases. We follow a correlated random-effects approach, and rely on additional layers of quantile regressions as a flexible tool to model conditional distributions. Conditions are given under which the model is nonparametrically identified in static or Markovian dynamic models. We develop a sequential method-of-moment approach for estimation, and compute the estimator using an iterative algorithm that exploits the computational simplicity of ordinary quantile regression in each iteration step. Finally, a Monte-Carlo exercise and an application to measure the effect of smoking during pregnancy on children's birthweights complete the paper.

JEL code: C23.
Keywords: Panel data, quantile regression, Expectation-Maximization.

*We thank Andrew Chesher, Bryan Graham, Roger Koenker, Konrad Menzel, Lars Nesheim, Yin Wei, and seminar participants at various places for comments. Raffaele Saggio provided excellent research assistance. Support from the European Research Council/ ERC grant agreement n^{0}263107 is gratefully acknowledged. All errors are our own.

1 Introduction

Nonlinear panel data models are central to applied research. However, despite some recent progress, it is fair to say that we are still short of answers for panel versions of many models commonly used in empirical work.[1] In this paper we focus on one particular nonlinear model for panel data: quantile regression.

Since Koenker and Bassett (1978), quantile regression has become a prominent methodology for examining the effects of explanatory variables across the entire outcome distribution. Extending the quantile regression approach to panel data has proven challenging, however, mostly because of the difficulty to handle individual-specific heterogeneity. Starting with Koenker (2004), most panel data approaches to date proceed in a quantile-by-quantile fashion, and include individual dummies as additional covariates in the quantile regression. As shown by some recent work, however, this *fixed-effects* approach faces special challenges when applied to quantile regression. Galvao, Kato and Montes-Rojas (2012) develop the large-N,T analysis of the fixed-effects quantile regression estimator, and show that it may suffer from large asymptotic biases. Rosen (2010) shows that the fixed-effects model for a single quantile is not point-identified.[2]

We depart from the previous literature by proposing a *random-effects* approach for quantile models from panel data. This approach treats individual unobserved heterogeneity as time-invariant missing data. To describe the model, let $i = 1, ..., N$ denote individual units, and let $t = 1, ..., T$ denote time periods. The random-effects quantile regression (REQR) model specifies the τ-specific conditional quantile of an outcome variable Y_{it}, given a sequence of strictly exogenous covariates $X_i = (X_{i1}', ..., X_{iT}')'$ and unobserved heterogeneity η_i, as follows:

$$Q\left(Y_{it} \mid X_i, \eta_i, \tau\right) = X_{it}'\beta\left(\tau\right) + \eta_i\gamma\left(\tau\right), \qquad \text{for all } \tau \in (0,1). \tag{1}$$

Note that η_i does not depend on the percentile value τ. Were data on η_i available, one could use a standard quantile regression package to recover the parameters $\beta\left(\tau\right)$ and $\gamma\left(\tau\right)$.

Model (1) specifies the conditional distribution of Y_{it} given X_{it} and η_i. In order to complete the model, we also specify the conditional distribution of η_i given the sequence of covariates X_i. For this purpose, we introduce an additional layer of quantile regression and

[1]See, for example, the survey by Arellano and Bonhomme (2011).
[2]Recent related contributions are Lamarche (2010), Galvao (2011), and Canay (2010).

specify the τ-th conditional quantile of η_i given covariates as follows:

$$Q\left(\eta_i \mid X_i, \tau\right) = X_i'\delta(\tau), \quad \text{for all } \tau \in (0,1). \tag{2}$$

This modelling allows for a flexible conditioning on strictly exogenous regressors—and on initial conditions in dynamic settings—that may also be of interest in other panel data models. Together, equations (1)-(2) provide a fully specified semiparametric model for the joint distribution of outcomes given the sequence of strictly exogenous covariates. The aim is then to recover the model's parameters: $\beta(\tau)$, $\gamma(\tau)$, and $\delta(\tau)$, for all τ.

We start by studying identification of the random-effects quantile regression model (1)-(2), under the assumption that outcomes in different periods are conditionally independent of each other given covariates and unobserved effects. To provide conditions under which the conditional distribution of Y_{it} given η_i and X_{it}, as well as the distribution of η_i given covariates, are both identified, we apply a result in Hu and Schennach (2008)—originally developed in the context of nonlinear measurement errors models—to our panel data context. Imposing some form of dynamic restrictions is necessary in order to separate out what part of the overall time variation is due to unobserved heterogeneity (Evdokimov, 2010, Arellano and Bonhomme, 2012). Moreover, building on Hu and Shum (2012), we show that a similar approach applies when Markovian restrictions, instead of full independence restrictions, are assumed to hold, thus extending the identification arguments to dynamic settings.

Our identification result for the REQR model is nonparametric. In particular, identification holds even if the conditional distribution of individual effects is left unrestricted. Recent research has emphasized the identification content of nonlinear panel data models with continuous outcomes (Bonhomme, 2012), as opposed to discrete outcomes models where parameters of interest are typically set-identified (Honoré and Tamer, 2006, Chernozhukov, Fernández-Val, Hahn and Newey, 2011). Pursuing this line of research, our analysis provides conditions for nonparametric identification of REQR in panels where the number of time periods T is fixed, possibly very short (e.g., $T = 3$). One of the required conditions to apply Hu and Schennach (2008)'s result is a *completeness* assumption. Although completeness is a high-level assumption, recent papers have provided primitive conditions in specific models, including a special case of model (1).[3]

[3]D'Haultfoeuille (2011), Andrews (2011), and Hu and Shiu (2012) provide primitive assumptions for completeness in specific models. See also Canay, Santos and Shaikh (2012) for a negative result on the power of tests of the completeness condition.

The main contribution of this paper is to provide an estimator of the parameters in the REQR model (1)-(2). To outline the intuition behind our estimation approach, it is useful to think of a standard quantile regression on an augmented sample of outcomes, covariates, and a (large) sequence of draws, say $\eta_i^{(m)}$ $(m = 1, ..., M)$ for each individual unit. We note that this augmented regression will deliver consistent estimates of $\beta(\tau)$ and $\gamma(\tau)$, provided $\eta_i^{(m)}$ are drawn from the *posterior* density of η_i given outcomes and covariates. This augmented quantile regression is infeasible, however, as the posterior distribution, which is proportional to the product of the random-effects distribution and the likelihood function, depends on the parameters to be estimated—in fact, on a continuum of parameters indexed by $\tau \in (0, 1)$.

The feasible estimator that we propose relies on an iterative approach. Given initial estimates of the model's parameters, we start by constructing model-implied estimates of the posterior distribution of η_i, using the fact that (1)-(2) is a complete semi-parametric model of the joint distribution of outcomes and individual effects. Then, given draws from the estimated posterior distribution, we update the quantile-specific parameters using simple quantile regressions. These two steps are iterated back and forth until convergence to a stationary distribution. Importantly, as weighted quantile regressions are linear programs, the algorithm preserves the computational simplicity of quantile regression within each iteration step.

Our estimation algorithm is related to several methods recently proposed in the literature. The sequential method-of-moments approach is related to the class of sequential estimators for finite mixture models developed by Arcidiacono and Jones (2003), as an extension of the classical Expectation-Maximization (EM) algorithm of Dempster, Laird and Rubin (1977).[4] Elashoff and Ryan (2004) present an algorithm for accommodating missing data in situations where a natural set of estimating equations exists for the complete data setting. Our algorithm is also related to, but different from, the EM algorithm with a Monte Carlo E-step (McLachlan and Krishnan, 2007). Geraci and Bottai (2007) considered a random-effects approach for a single quantile assuming that the outcome variable is distributed as an asymmetric Laplace (for given τ) conditional on covariates and individual effects, and used a Monte Carlo EM algorithm for the computation of the Maximum Likelihood estimator. In addition, our approach is related to Abrevaya and Dahl (2008), who considered a correlated random-effects model to study the effects of smoking and prenatal care on birthweight. Their

[4]See also Arcidiacono and Miller (2011) and Bonhomme and Robin (2009) for related approaches.

approach mimics control function approaches used in linear panel models.

Our analysis is most closely related to Wei and Carroll (2009), who proposed a consistent estimation method for cross-sectional linear quantile regression subject to covariate measurement error. In particular, we rely on the approach in Wei and Carroll to deal with the continuum of model parameters indexed by $\tau \in (0,1)$. As keeping track of all parameters in the algorithm is not feasible, we build on their insight and use interpolating splines to combine the quantile-specific parameters in (1)-(2) into a complete likelihood function that depends on a finite number of parameters. Our proof of consistency—in a panel data asymptotics where N tends to infinity and T is kept fixed—also builds on theirs. As the sample size increases, the number of knots, and hence the accuracy of the spline approximation, increase as well. A key difference with Wei and Carroll is that, in our setup, the conditional distribution of individual effects is unknown, and needs to be estimated along with the other parameters of the model.

An attractive feature of the REQR approach is that it may be generalized to deal with a large class of models, thus providing a general estimation approach for panel data applications. A first extension is to allow for multiple individual effects, i.e., for vector-valued η_i in model (1)-(2). To do so, we model conditional distributions in a triangular fashion. For example, with two individual effects we model the distribution of η_{i2} given η_{i1} and X_i, and that of η_{i1} given X_i, both using linear quantile regression models. Though not invariant to ordering choice, this approach offers a flexible tool to model conditional multivariate distributions. A second important class of extensions is to allow for dynamics. We show how to extend the setup to autoregressive models, models with general predetermined regressors, and models with autocorrelated errors. These extensions can be accommodated using slight modifications of the baseline algorithm. In particular, models with predetermined regressors are handled using additional layers of quantile regressions to model the feedback process.

We present some preliminary evidence on simulated data that suggests good finite-sample performance. In addition, we apply the REQR estimator to assess the effect of smoking during pregnancy on a child's birth outcomes. Following Abrevaya (2006), we allow for mother-specific fixed-effects in estimation. Both nonlinearities and unobserved heterogeneity are thus allowed for by our panel data quantile regression estimator. We find that, while allowing for time-invariant mother-specific effects decreases the magnitude of the negative coefficient of smoking, the latter remains sizable, especially at low birthweights.

The rest of the paper is as follows. In Section 2 we present the model and provide conditions for nonparametric identification. In Section 3 we describe the estimation algorithm, and we study its asymptotic properties in Section 4. In Section 5 we discuss implementation issues, and in Section 6 we present various extensions to dynamic models. In Section 7 we show numerical evidence on the performance of the estimator, and we apply the method to estimate the effect of smoking on children's birthweight. Lastly, we conclude in Section 8.

2 Model and identification

In this section and the next we focus on the static version of the random-effects quantile regression (REQR) model. Section 6 will consider various extensions to dynamic models. We start by presenting the model along with several examples, and then provide conditions for nonparametric identification.

2.1 Model

Let $Y_i = (Y_{i1}, ..., Y_{iT})'$ denote a sequence of T scalar outcomes for individual i, and let $X_i = (X_{i1}', ..., X_{iT}')'$ denote a sequence of strictly exogenous regressors, which may contain a constant. In addition, let η_i denote a q-dimensional vector of individual-specific effects, and let U_{it} denote a scalar error term. The model specifies the conditional quantile response function of Y_{it} given X_{it} and η_i as follows:

$$Y_{it} = Q_Y\left(X_{it}, \eta_i, U_{it}\right) \qquad i = 1, ..., N, \quad t = 1, ..., T. \tag{3}$$

We make the following assumptions.

Assumption 1 *(outcomes)*

(i) U_{it} *follows a standard uniform distribution conditional on* X_i *and* η_i*.*

(ii) $\tau \mapsto Q\left(x, \eta, \tau\right)$ *is strictly increasing on* $(0, 1)$*, almost surely in* (x, η)*.*

(iii) U_{it} *is independent of* U_{is} *for each* $t \neq s$ *conditional on* X_i *and* η_i*.*

Assumption 1 (i) contains two parts. First, U_{it} is assumed independent of the full sequence $X_{i1}, ..., X_{IT}$, and independent of individual effects. This assumption of strict exogeneity rules out predetermined or endogenous covariates. Second, the marginal distribution of U_{it} is normalized to be uniform on the unit interval. Part (ii) guarantees that outcomes have absolutely continuous distributions. Together, parts (i) and (ii) imply that, for any given $\tau \in (0,1)$, $Q_Y(X_{it}, \eta_i, \tau)$ is the τ-conditional quantile of Y_{it} given X_i and η_i.[5]

Lastly, Assumption 1 (iii) imposes independence restrictions on the process $(U_{i1}, ..., U_{iT})$. Restricting the dynamics of error variables U_{it} is essential to separate the time-varying unobserved shocks from the time-invariant unobserved individual effects η_i. Part (iii) defines the static version of the model, by assuming that U_{it} are i.i.d. over time. In Section 6 we will develop various extensions of the model that allow for dynamic effects and non-strictly exogenous regressors.

In addition, the model specifies the conditional quantile response function of η_i given X_i as follows:

$$\eta_{it} = Q_\eta(X_i, V_i) \qquad i = 1, ..., N. \tag{4}$$

Note that, provided η_i is continuously distributed given X_i and Assumption 2 below holds, equation (4) is a representation that comes without loss of generality, corresponding to a fully unrestricted correlated random-effects specification.

Assumption 2 *(individual effects)*

(i) V_i *follows a standard uniform distribution conditional on* X_i.

(ii) $\tau \mapsto Q_\eta(x, \tau)$ *is strictly increasing on* $(0,1)$, *almost surely in* x.

We now describe four examples in turn.

Example 1 (Location-scale). A simple example is the panel generalization of the location-scale model (He, 1997):

$$Y_{it} = X_{it}'\beta + \eta_i + (X_{it}'\gamma + \mu\eta_i)\,\varepsilon_{it}, \tag{5}$$

where ε_{it} are i.i.d. across periods, and independent of all regressors and individual effects.[6]

[5]Indeed we have, using Assumption 1 (i) and (ii):

$$\begin{aligned}\Pr\left(Y_{it} \leq Q_Y(X_{it}, \eta_i, \tau)\,|X_i, \eta_i\right) &= \Pr\left(Q_Y(X_{it}, \eta_i, U_{it}) \leq Q_Y(X_{it}, \eta_i, \tau)\,|X_i, \eta_i\right) \\ &= \Pr\left(U_{it} \leq \tau | X_i, \eta_i\right) = \tau.\end{aligned}$$

[6]Note that a generalization of (5) that allows for two-dimensional individual effects—as in Example 4 below—is:

$$Y_{it} = X_{it}'\beta + \eta_{i1} + (X_{it}'\gamma + \eta_{i2})\,\varepsilon_{it}.$$

Denoting $U_{it} = F(\varepsilon_{it})$, where F is the cdf of ε_{it}, the conditional quantiles of Y_{it} are given by:

$$Q_Y(X_{it}, \eta_i, \tau) = X_{it}'\beta + \eta_i + (X_{it}'\gamma + \mu\eta_i) F^{-1}(\tau), \quad \tau \in (0,1).$$

Example 2 (Linear quantiles). In our second example we focus on the following linear quantile specification with scalar η_i:

$$Y_{it} = X_{it}'\beta(U_{it}) + \eta_i\gamma(U_{it}). \tag{6}$$

Note that, given Assumption 1 (*i*) and (*ii*), the conditional quantiles of Y_{it} are given by (1). Note also that the location-scale model (5) is a special case of (6).

Model (6) is a panel data generalization of the classical linear quantile model of Koenker and Bassett (1978). Were we to observe the individual effects η_i along with the covariates X_{it}, it would be reasonable to postulate a model of this form. It is instructive to compare model (6) with the following more general but different type of model:

$$Y_{it} = X_{it}'\beta(U_{it}) + \eta_i(U_{it}),$$

where $\eta_i(\tau)$ is an individual-specific nonparametric function of τ. Koenker (2004) and subsequent fixed-effects approaches considered this more general model. Unlike the REQR model, the presence of the process $\eta_i(\tau)$ introduces an element of nonparametric functional heterogeneity in the conditional distribution of Y_{it}.

In order to complete model (6)—or, indeed, model (5)—one may use another linear quantile specification for the conditional distribution of individual effects:

$$\eta_i = X_i'\delta(V_i). \tag{7}$$

Given Assumption 2, the conditional quantiles of η_i are then given by (2).

Model (7) corresponds to a *correlated random-effects* approach. However, it is more flexible than alternative specifications in the literature. A commonly used specification is (Chamberlain, 1984):

$$\eta_i = X_i'\mu + \sigma\varepsilon_i, \quad \varepsilon_i | X_i \sim \mathcal{N}(0,1). \tag{8}$$

For example, in contrast with (8), model (7) is fully nonparametric in the absence of covariates—i.e., when an *independent* random-effects specification is assumed. It is worth pointing out that model (7) might also be of interest in other nonlinear panel data models, where the outcome equation does not follow a linear quantile model. We will return to this point in the conclusion.

We will refer to model (6)-(7) as the static linear REQR model. In the next section we develop an estimator for this model, and derive its asymptotic properties.

Example 3 (Multiple individual effects). The linear REQR model may easily be modified to allow for more general interactions between observables and unobservables. To motivate this extension, note that, in Example 2, the *quantile marginal effect* for individual i associated with a marginal increase in covariates is the same for all units, as:

$$\frac{\partial}{\partial X_{it}} Q_Y\left(X_{it}, \eta_i, \tau\right) = \beta(\tau).$$

A random coefficients generalization that allows for heterogeneous quantile effects is:

$$Q_Y\left(X_{it}, \eta_i, \tau\right) = X_{it}'\beta(\tau) + \gamma_1(\tau)\eta_{i1} + X_{it}'\gamma_2(\tau)\eta_{i2}, \tag{9}$$

where $\eta_i = (\eta_{i1}, \eta_{i2})'$ is bivariate. The quantile marginal effect for individual i is then:

$$\frac{\partial}{\partial X_{it}} Q_Y\left(X_{it}, \eta_i, \tau\right) = \beta(\tau) + \gamma_2(\tau)\eta_{i2}.$$

In order to extend (7) to the case with bivariate unobserved heterogeneity, it is convenient to assume a triangular structure such as:

$$\begin{cases} \eta_{i1} = X_i'\delta_{11}\left(V_{i1}\right), \\ \eta_{i2} = \eta_{i1}\delta_{21}\left(V_{i2}\right) + X_i'\delta_{22}\left(V_{i2}\right), \end{cases} \tag{10}$$

where V_{i1} and V_{i2} follow independent standard uniform distributions. Though not invariant to permutation of (η_{i1}, η_{i2})—except if fully nonparametric—model (10) provides a flexible specification for the bivariate conditional distribution of (η_{i1}, η_{i2}) given X_i.[7]

Example 4 (Smooth coefficients) Our last example is the following semiparametric generalization with a richer dependence on individual effects:

$$Y_{it} = X_{it}'\beta\left(U_{it}, \eta_i\right) + \gamma\left(U_{it}, \eta_i\right) \tag{11}$$

where $\beta\left(\tau, .\right)$ and $\gamma\left(\tau, .\right)$ are smooth coefficient functions of the random effects.

[7]It is worth pointing out that quantiles appear not to generalize easily to the multivariate case. Multivariate quantile regression is still an open research area.

For any given point η^*, a local polynomial quantile method uses the approximations

$$\beta(\tau, \eta_i) \approx \sum_{j=0}^{q} b_{\tau j}(\eta^*)(\eta_i - \eta^*)^j, \quad \text{and} \quad \gamma(\tau, \eta_i) \approx \sum_{j=0}^{q} g_{\tau j}(\eta^*)(\eta_i - \eta^*)^j,$$

where $\beta(\tau, \eta^*) = b_{\tau 0}(\eta^*)$ and $\gamma(\tau, \eta^*) = g_{\tau 0}(\eta^*)$. Our approach may be adapted to this setup, using locally weighted check function for estimation (Chaudhuri, 1991, Cai and Xu, 2008). An attractive feature of model (11) is that it exhibits similar flexibility in transitory and permanent unobservable components.

As a related approach, one could use semiparametric methods—both in the X and τ dimensions—to generalize model (7), and flexibly model the conditional quantile function of individual effects given covariates. See Belloni, Chernozhukov and Fernández-Val (2012) for an approach based on series expansions, and Qu and Yoon (2011) for a kernel-based approach. We leave the asymptotic studies of these generalizations to future work.

2.2 Identification

In this section we study nonparametric identification in model (3)-(4). We start with the case where there is a single scalar individual effect (i.e., $q = \dim \eta_i = 1$), and we set $T = 3$.

Under conditional independence over time— Assumption 1 (iii)—we have, for all y_1, y_2, y_3, $x = (x_1', x_2', x_3')'$, and η:

$$f_{Y_1,Y_2,Y_3|\eta,X}(y_1, y_2, y_3 \mid \eta, x) = f_{Y_1|\eta,X}(y_1 \mid \eta, x) f_{Y_2|\eta,X}(y_2 \mid \eta, x) f_{Y_3|\eta,X}(y_3 \mid \eta, x). \tag{12}$$

Hence the data distribution function relates to the densities of interest as follows:

$$\begin{aligned} f_{Y_1,Y_2,Y_3|X}(y_1, y_2, y_3 \mid x) &= \int f_{Y_1|\eta,X}(y_1 \mid \eta, x) f_{Y_2|\eta,X}(y_2 \mid \eta, x) f_{Y_3|\eta,X}(y_3 \mid \eta, x) \\ &\qquad \times f_{\eta|X}(\eta \mid x)\, d\eta. \end{aligned} \tag{13}$$

The goal is the identification of $f_{Y_1|\eta,X}$, $f_{Y_2|\eta,X}$, $f_{Y_3|\eta,X}$ and $f_{\eta|X}$ given knowledge of $f_{Y_1,Y_2,Y_3|X}$.

The setting of equation (13) is formally equivalent (conditional on x) to the instrumental variables setup of Hu and Schennach (2008), for nonclassical nonlinear errors-in-variables models. Specifically, according to Hu and Schennach's terminology Y_3 would be the outcome variable, Y_2 would be the mismeasured regressor, Y_1 would be the instrumental variable, and η would be the latent, error-free regressor. We closely rely on their analysis and make the following additional assumptions.

Assumption 3 *(identification) Almost surely in covariate values x:*

(i) The joint density $f_{Y_1,Y_2,Y_3,\eta|X}(\cdot,\cdot,\cdot,\cdot|x)$ is bounded, as well as all its joint and marginal densities.

(ii) For all $\eta_1 \neq \eta_2$: $\Pr\left[f_{Y_3|\eta,X}(Y_3|\eta_1,x) \neq f_{Y_3|\eta,X}(Y_3|\eta_2,x)|X=x\right] > 0$.

(iii) There exists a known functional Γ_x such that $\Gamma_x(f_{Y_2|\eta,X}(\cdot|\eta,x)) = \eta$.

(iv) The linear operators $L_{Y_2|\eta,x}$ and $L_{Y_1|Y_2,x}$, associated with the conditional densities $f_{Y_2|\eta,X}(\cdot \mid \cdot,x)$ and $f_{Y_1|Y_2,X}(\cdot \mid \cdot,x)$, respectively, are injective.

Part (i) in Assumption 3 requires that all densities under consideration be bounded. This imposes mild restrictions on the model's parameters.[8] Part (ii) requires that $f_{Y_3|\eta,X}$ be nonidentical at different values of η. It is easy to see that this assumption will be satisfied if, for some τ in small open neighborhood:

$$Q_Y\left(Y_3|\eta=\eta_1, X=x,\tau\right) \neq Q_Y\left(Y_3|\eta=\eta_2, X=x,\tau\right).$$

In the linear REQR model of Example 2, this holds if $\gamma(\tau) \neq 0$ for τ in some neighborhood.

Part (iii) is a normalization assumption, which imposes a centered measure of location on $f_{Y_2|\eta,X}$. In fact, in order to apply the identification theorem in Hu and Schennach (2008), it is not necessary that Γ_x be known. If instead Γ_x is a known *function* of the data distribution, their argument goes through. For example, in the linear REQR model of Example 2, one convenient normalization is obtained by noting that:

$$\begin{aligned}
\mathbb{E}\left(Y_{it}|\eta_i, X_{it}\right) &= X_{it}'\left[\int_0^1 \beta\left(\tau\right)d\tau\right] + \eta_i\left[\int_0^1 \gamma\left(\tau\right)d\tau\right] \\
&\equiv \widetilde{X}_{it}'\overline{\beta}_1 + \overline{\beta}_0 + \eta_i\overline{\gamma},
\end{aligned}$$

where $\overline{\beta}_0 = \int_0^1 \beta_0(\tau)d\tau$ corresponds to the coefficient of the constant in $X_{it} = (\widetilde{X}_{it}', 1)'$. Now, if $\widetilde{X}_{it}$ varies over time, $\overline{\beta}_1$ is a known function of the data distribution, simply given by the within-group estimand. In this case one may thus take:

$$\Gamma_x(g) = \int yg(y)dy - \widetilde{x}_2'\overline{\beta}_1,$$

and note that the following normalization implies Assumption 3 (iii):

$$\overline{\beta}_0 = \int_0^1 \beta_0\left(\tau\right)d\tau = 0, \quad \text{and} \quad \overline{\gamma} = \int_0^1 \gamma\left(\tau\right)d\tau = 1. \tag{14}$$

[8]For example, in the linear REQR model of Example 2, we need in particular strict monotonicity of quantile functions; that is: $x_t'\nabla\beta(\tau) + \eta\nabla\gamma(\tau) \geq \underline{c} > 0$, where $\nabla\xi(\tau)$ denotes the first derivative of $\xi(\cdot)$ evaluated at τ.

We will use normalization (14) in the simulation exercises and the empirical application.

The last part in Assumption 3—part (*iv*)—is an *injectivity* condition. The operator $L_{Y_2|\eta,X}$ is defined as $[L_{Y_2|\eta,x}h](y_2) = \int f_{Y_2|\eta,X}(y_2|\eta,x)h(\eta)d\eta$, for all bounded functions h. $L_{Y_2|\eta,x}$ is injective if $L_{Y_2|\eta,x}h = 0 \Rightarrow h = 0$. As pointed out by Hu and Schennach (2008), injectivity is closely related to *completeness* conditions commonly assumed in the literature on nonparametric instrumental variable estimation. Like completeness, injectivity is a high-level, non-transparent condition; see for example Canay *et al.* (2012).

Several recent papers provide explicit conditions for completeness or injectivity in specific models. Andrews (2011) constructs classes of distributions that are L^2-complete and boundedly complete. D'Haultfoeuille (2011) provides primitive conditions for completeness in a linear model with homoskedastic errors. The recent work by Hu and Shiu (2012) considers a more general class of models. In particular, they provide conditions for completeness in location-scale models. Their results apply to the location-scale quantile model (5). In this case, conditions that guarantee that $L_{Y_2|\eta,x}$ is injective involve the tail properties of the conditional density of Y_2 given η (and x) and its characteristic function.[9] Providing primitive conditions for injectivity/completeness in more general models, such as the linear REQR model of Example 2, is an interesting question but exceeds the scope of this paper.

We then have the following result, which is a direct application of the identification theorem in Hu and Schennach (2008). For completeness, a sketch of the identification argument is given in Appendix C.

Proposition 1 *(Hu and Schennach, 2008)*

Let Assumptions 1, 2, and 3 hold. Then all conditional densities $f_{Y_1|\eta,X}$, $f_{Y_2|\eta,X}$, $f_{Y_3|\eta,X}$, and $f_{\eta|X}$, are nonparametrically identified.

Models with multiple effects. The identification result extends to models with multiple individual effects $\eta_i \in \mathbb{R}^q$ with $q > 1$, taking a larger $T > 3$. For example, with $T = 5$ it is possible to apply Hu and Schennach (2008)'s identification theorem to a bivariate η using (Y_1, Y_2) instead of Y_1, (Y_3, Y_4) instead of Y_2, and Y_5 instead of Y_3. Provided injectivity conditions hold, nonparametric identification follows from similar arguments as above. Note that, in this case, the overidentifying restrictions suggest the possibility of identification with

[9] See Lemma 4 in Hu and Shiu (2012).

$T = 4$, although proving this conjecture would require using techniques other than the ones used by Hu and Schennach.[10]

3 REQR estimation

This section considers estimation in the static model (6)-(7). We start by describing the moment restrictions that our estimator exploits, and then present the sequential estimator. In the next two sections we will study the asymptotic properties of the estimator and discuss implementation issues in turn.

3.1 Moment restrictions

We start by setting the notation. Let $\theta(\cdot) = \left(\beta(\cdot)', \gamma(\cdot)\right)'$. We denote as $f(\eta \mid y, x)$ the posterior density of the individual effects:

$$f(\eta \mid y, x; \theta(\cdot), \delta(\cdot)) = \frac{\prod_{t=1}^{T} f(y_t \mid x_t, \eta; \theta(\cdot)) f(\eta \mid x; \delta(\cdot))}{\int \prod_{t=1}^{T} f(y_t \mid x_t, \widetilde{\eta}; \theta(\cdot)) f(\widetilde{\eta} \mid x; \delta(\cdot)) d\widetilde{\eta}}, \tag{15}$$

where we have used the conditional independence assumption 1 (*iii*), and where we have explicitly indicated the dependence of the various densities on model parameters.

Let $\psi_\tau(u) = \tau - \mathbf{1}\{u < 0\}$. Note that ψ_τ is the first derivative (outside the origin) of the *check* function ρ_τ, which is familiar from the quantile regression literature (Koenker and Basset, 1978): $\rho_\tau(u) = (\tau - \mathbf{1}\{u < 0\}) u$, and $\psi_\tau(u) = \nabla \rho(u)$. Let also $W_{it}(\eta) = (X_{it}', \eta)'$.

In order to derive the main moment restrictions, we start by noting that, for all $\tau \in (0, 1)$, the following infeasible moment restrictions hold, as a direct implication of Assumptions 1 and 2:

$$\mathbb{E}\left[\sum_{t=1}^{T} W_{it}(\eta_i) \psi_\tau\left(Y_{it} - W_{it}(\eta_i)' \theta(\tau)\right)\right] = 0, \tag{16}$$

and:

$$\mathbb{E}\left[X_i \psi_\tau\left(\eta_i - X_i' \delta(\tau)\right)\right] = 0. \tag{17}$$

Indeed, (16) is the first-order condition associated with the infeasible population quantile regression of Y_{it} on X_{it} and η_i. Similarly, (17) corresponds to the infeasible quantile regression of η_i on X_i.

[10] A similar comment applies to the linear quantile model (6). In this case the overidentifying restrictions suggest that the model with $T = 2$ might be identified. Note, however, that even if identification could be shown, it would fundamentally rely on the linear quantile specification, as opposed to the nonparametric identification result of Proposition 1.

Applying the law of iterated expectations to (16) and (17), respectively, we obtain our main moment restrictions, for all $\tau \in (0,1)$:

$$\mathbb{E}\left[\int \left(\sum_{t=1}^{T} W_{it}(\eta)\,\psi_\tau\left(Y_{it} - W_{it}(\eta)'\theta(\tau)\right)\right) f\left(\eta \mid Y_i, X_i; \theta(\cdot), \delta(\cdot)\right) d\eta\right] = 0, \tag{18}$$

and:

$$\mathbb{E}\left[\int \left(X_i\psi_\tau\left(\eta - X_i'\delta(\tau)\right)\right) f\left(\eta \mid Y_i, X_i; \theta(\cdot), \delta(\cdot)\right) d\eta\right] \;=\; 0. \tag{19}$$

It follows from (18)-(19) that, if the posterior density of the individual effects were known, then estimating the model's parameters could be done using two simple linear quantile regressions, weighted by the posterior density. However, as the notation makes clear, the posterior density in (15) depends on the entire processes $\theta(\cdot)$ and $\delta(\cdot)$. Specifically we have, provided the conditional densities of outcomes and individual effects be absolutely continuous:

$$f\left(y_t \mid x_t, \eta; \theta(\cdot)\right) = \lim_{\epsilon \to 0} \frac{\epsilon}{w_t(\eta)'\left[\theta(u_t + \epsilon) - \theta(u_t)\right]}, \tag{20}$$

and:

$$f\left(\eta \mid x; \delta(\cdot)\right) = \lim_{\epsilon \to 0} \frac{\epsilon}{x'\left[\delta(v + \epsilon) - \delta(v)\right]}, \tag{21}$$

where u_t and v are defined by:

$$w_t(\eta)'\theta(u_t) = y_t, \qquad \text{and:} \qquad x'\delta(v) = \eta.$$

Equations (20)-(21) come from the fact that the density of a random variable and the derivative of its quantile function are the inverse of each other.

The dependence of the posterior density on the entire set of model parameters makes it impossible to recover $\theta(\tau)$ and $\delta(\tau)$ in (18)-(19) in a τ-by-τ fashion. The main idea of the algorithm that we present next is to circumvent this difficulty by iterating back-and-forth between computation of the posterior density, and computation of the model's parameters given the posterior density. The latter is easy to do as it is based on weighted quantile regressions. Similar ideas have been used in the literature (e.g., Arcidiacono and Jones, 2003). However, one additional difficulty in our case is that the posterior density depends on a continuum of parameters. In order to develop a practical approach, we now introduce a finite-dimensional, tractable approximating model.

3.2 Approximating model

Building on Wei and Carroll (2009), we approximate $\theta(\cdot)$ and $\delta(\cdot)$ using *splines*, with L knots $0 < \tau_1 < \tau_2 < ... < \tau_L < 1$. A practical possibility is to use piecewise-linear splines as in Wei and Carroll, but other choices are possible, such as cubic splines or shape-preserving B-splines. When using interpolating splines, the approximation argument requires suitable smoothness assumptions on $\theta(\tau)$ and $\delta(\tau)$ as functions of $\tau \in (0,1)$. In the consistency proof in the next section we will let L increase with the sample size at an appropriate rate.

Let us define $\xi = (\xi_A', \xi_B')'$, where:

$$\xi_A = \left(\theta(\tau_1)', \theta(\tau_2)', ..., \theta(\tau_L)'\right)', \qquad \text{and} \qquad \xi_B = \left(\delta(\tau_1)', \delta(\tau_2)', ..., \delta(\tau_L)'\right)'.$$

The approximating model depends on the finite-dimensional parameter vector ξ that is used to construct interpolating splines. The associated likelihood functions and random-effects density are then denoted as $f(y_t \mid x_t, \eta; \xi_A)$ and $f(\eta \mid x; \xi_B)$, respectively, and the implied posterior density is:

$$f(\eta \mid y, x; \xi) = \frac{\prod_{t=1}^T f(y_t \mid x_t, \eta; \xi_A) f(\eta \mid x; \xi_B)}{\int \prod_{t=1}^T f(y_t \mid x_t, \widetilde{\eta}; \xi_A) f(\widetilde{\eta} \mid x; \xi_B) d\widetilde{\eta}}. \tag{22}$$

The approximating densities take particularly simple forms when using piecewise-linear splines; see Section 5.1 below.

The moment restrictions of the approximating model are then, for all $\ell = 1, ..., L$:

$$\mathbb{E}\left[\int \left(\sum_{t=1}^T W_{it}(\eta)\psi_{\tau_\ell}\left(Y_{it} - W_{it}(\eta)'\theta(\tau_\ell)\right)\right) f(\eta \mid Y_i, X_i; \xi) d\eta\right] = 0, \tag{23}$$

and:

$$\mathbb{E}\left[\int \left(X_i \psi_{\tau_\ell}(\eta - X_i'\delta(\tau_\ell))\right) f(\eta \mid Y_i, X_i; \xi) d\eta\right] \;=\; 0. \tag{24}$$

3.3 Sequential estimator

Let (y_i, x_i), $i = 1, ..., N$, be an i.i.d. sample. Our estimator is the solution to the following sample fixed-point problem, for $\ell = 1, ..., L$:

$$\widehat{\theta}(\tau_\ell) \;=\; \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^N \int \left(\sum_{t=1}^T \rho_{\tau_\ell}\left(y_{it} - w_{it}(\eta)'\theta\right)\right) f\left(\eta \mid y_i, x_i; \widehat{\xi}\right) d\eta, \tag{25}$$

$$\widehat{\delta}(\tau_\ell) \;=\; \underset{\delta}{\operatorname{argmin}} \sum_{i=1}^N \int \rho_{\tau_\ell}(\eta - x_i'\delta) f\left(\eta \mid y_i, x_i; \widehat{\xi}\right) d\eta, \tag{26}$$

where $\rho_\tau(\cdot)$ is the check function, and where $f(\eta \mid y_i, x_i; \xi)$ is given by (22). Note that the first-order conditions of (25)-(26) are the sample analogs of the moment restrictions (23)-(24) of the approximating model.

To solve the fixed-point problem (25)-(26) we proceed in an iterative fashion. Starting with initial parameter values $\widehat{\xi}^{(0)}$ we iterate the following two steps until numerical convergence:

1. Compute the posterior density:

$$\widehat{f}_i^{(s)}(\eta) = f\left(\eta \mid y_i, x_i; \widehat{\xi}^{(s)}\right). \tag{27}$$

2. Solve, for $\ell = 1, ..., L$:

$$\widehat{\theta}(\tau_\ell)^{(s+1)} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} \int \left(\sum_{t=1}^{T} \rho_{\tau_\ell}\left(y_{it} - w_{it}(\eta)'\theta\right) \right) \widehat{f}_i^{(s)}(\eta) d\eta, \tag{28}$$

$$\widehat{\delta}(\tau_\ell)^{(s+1)} = \underset{\delta}{\operatorname{argmin}} \sum_{i=1}^{N} \int \rho_{\tau_\ell}(\eta - x_i'\delta) \widehat{f}_i^{(s)}(\eta) d\eta. \tag{29}$$

This sequential method-of-moment method is related to, but different from, the standard EM algorithm (Dempster *et al.*, 1977). Like in EM, the algorithm iterates back-and-forth between computation of the posterior density of the individual effects ("E"-step) and computation of the parameters given the posterior density ("M"-step). Unlike in EM, however, in the "M"-step of our algorithm—i.e., in equations (28)-(29)—estimation is not based on a likelihood function, but on the *check* function of quantile regression.

Proceeding in this way has two major computational advantages compared to maximizing the full likelihood of the approximating model. Firstly, as opposed to the likelihood function, which is a complicated function of all quantile regression coefficients, the problem in (28)-(29) nicely decomposes into L different τ_ℓ-specific subproblems. Secondly, using the check function yields a globally convex objective function in each "M"-step.

Note, however, that two features of the standard EM algorithm differ in our sequential method-of-moment method. First, as our algorithm is not likelihood-based, the resulting estimator will not be efficient in general.[11] Secondly, whereas conditions for numerical convergence of ordinary EM are available in the literature (e.g., Wu, 1983), formal proofs of

[11] This loss of efficiency relative to maximum likelihood is similar to the one documented in Arcidiacono and Jones (2003), for example.

convergence of sequential algorithms such as ours seem difficult to establish.[12] Lastly, note that, to implement the estimation algorithm, one needs to compute the integrals that appear in (27)-(28)-(29). We develop a stochastic version of the algorithm in Section 5.2.

4 Asymptotic properties

In this section we study the consistency and asymptotic distribution of the REQR estimator for fixed T, as N tends to infinity.

4.1 Consistency

To establish consistency of the REQR estimator we start by setting the notation. We let $\xi(\tau) = (\theta(\tau)', \delta(\tau)')'$ be a $p \times 1$ vector for all $\tau \in (0,1)$, and we let $\xi(\cdot) : (0,1) \to \mathbb{R}^p$ be the associated function. We rewrite the population moment restrictions (18)-(19) as:

$$\mathbb{E}\left[\Psi_i\left(\xi^0(\cdot), \tau\right)\right] \quad = \quad 0, \quad \tau \in (0,1), \tag{30}$$

where Ψ_i is $p \times 1$, and where $\xi^0(\cdot)$ denote the true value of the function $\xi(\cdot)$.

Equation (30) defines a continuum of moment restrictions on $\xi^0(\cdot)$. The next assumption defines the set of population functions $\xi^0(\cdot)$, for some constants $\bar{c} > 0$ and $\underline{c} > 0$. We will denote as $\mathcal{H}$ the set of functions that satisfy Assumption 4.

Assumption 4 *(population parameters)*

(i) $\xi^0(\cdot) : (0,1) \to \mathbb{R}^p$ *is differentiable. Moreover:*

$$\sup_{\tau \in (0,1)} \left\|\xi^0(\tau)\right\| + \sup_{\tau \in (0,1)} \left\|\nabla\xi^0(\tau)\right\| + \sup_{(\tau_1,\tau_2)\in(0,1)^2, \tau_1 \neq \tau_2} \frac{\left\|\nabla\xi^0(\tau_2) - \nabla\xi^0(\tau_1)\right\|}{|\tau_2 - \tau_1|} \leq \bar{c},$$

where the first derivatives in $\nabla\xi^0(\tau)$ *are component-wise.*

(ii) $\int_0^1 \beta_0^0(\tau)d\tau = 0$, *and* $\int_0^1 \gamma^0(\tau)d\tau = 1$.

(iii) $\inf_{\tau\in(0,1)} |\gamma^0(\tau)| \geq \underline{c}$.

(iv) For all y, x, η, t, *and all* $\tau_1 < \tau_2$:

$$\frac{w_t(\eta)'\left[\theta^0(\tau_2) - \theta^0(\tau_1)\right]}{(\tau_2 - \tau_1)} \geq \underline{c} \quad \text{ and } \quad \frac{x'\left[\delta^0(\tau_2) - \delta^0(\tau_1)\right]}{(\tau_2 - \tau_1)} \geq \underline{c}.$$

[12] Our algorithm belongs to the class of "EM algorithms for estimating equations" studied by Elashoff and Ryan (2004). These authors provide conditions for numerical convergence, while acknowledging that verifying these conditions in practice may be difficult.

Assumption 4 (i) imposes smoothness restrictions on $\xi^0(\cdot)$. Specifically, it requires that $\xi^0(\cdot)$ belong to an Hölder ball. These spaces of functions are commonly used in semiparametric models (e.g., Ai and Chen, 2003). Part (ii) requires that $\theta^0(\cdot)$ satisfy the mean normalizations (14). Part (iii) requires that $\gamma^0(\tau)$ be bounded from below. This will guarantee that the moment functions $\Psi_i\left(\xi(\cdot),\tau\right)$ are continuous in both arguments, and allow us to rely on well established techniques in the consistency proof. Lastly, part (iv) requires that the implied conditional densities of Y_{it} given η_i, X_{it}, and that of η_i given X_i, be bounded from above, uniformly in their arguments. This last requirement imposes a strict monotonicity condition on quantile functions.

In the next assumption we define the splines. We consider piecewise-linear splines, but the analysis would similarly apply to other spline families. Linear splines are a convenient choice for implementation, as we will see in Section 5.2.

Assumption 5 *(splines) Given an $L\geq 1$, let $\mathcal{H}^L$ be the space of functions that satisfy the following conditions.*

(i) $\xi(\cdot):(0,1)\rightarrow\mathbb{R}^p$ is continuous, piecewise-linear on (τ_1,τ_L) with knots $\tau_1,\tau_2,...,\tau_L$.[13]

(ii) $\sup_\ell \|\xi(\tau_\ell)\| + \sup_{\ell_1<\ell_2} \frac{\|\xi(\tau_{\ell_2})-\xi(\tau_{\ell_1})\|}{\tau_{\ell_2}-\tau_{\ell_1}} + \sup_{\ell_1<\ell_2<\ell_3} \frac{\left\|\frac{\xi(\tau_{\ell_3})-\xi(\tau_{\ell_2})}{\tau_{\ell_3}-\tau_{\ell_2}}-\frac{\xi(\tau_{\ell_2})-\xi(\tau_{\ell_1})}{\tau_{\ell_2}-\tau_{\ell_1}}\right\|}{\tau_{\ell_3}-\tau_{\ell_1}}\leq\overline{c}$.

(iii) $(\beta_0(\tau_1)+...+\beta_0(\tau_L))/L=0$, and $(\gamma(\tau_1)+...+\gamma(\tau_L))/L=1$.

(iv) $\min_{\ell\in\{1,...,L\}}|\gamma(\tau_\ell)|\geq\underline{c}$.

(v) For all y,x,η,t, and all $\ell_1<\ell_2$:

$$\frac{w_t(\eta)'\left[\theta(\tau_{\ell_2})-\theta(\tau_{\ell_1})\right]}{(\tau_{\ell_2}-\tau_{\ell_1})}\geq\underline{c}\quad \text{and}\quad \frac{x'\left[\delta(\tau_{\ell_2})-\delta(\tau_{\ell_1})\right]}{(\tau_{\ell_2}-\tau_{\ell_1})}\geq\underline{c}.$$

(vi) For $\tau\in(0,\tau_1)$, $\beta_1(\tau)=\beta_1(\tau_1)$, $\gamma(\tau)=\gamma(\tau_1)$, and the constant coefficient is $\beta_0(\tau)=\beta_0(\tau_1)+G(\tau)$, where $G(\cdot)$ is a known strictly increasing function such that $G(\tau_1)=0$. For $\tau\in(\tau_L,1)$, $\beta_1(\tau)=\beta_1(\tau_L)$, $\gamma(\tau)=\gamma(\tau_L)$, and $\beta_0(\tau)=\beta_0(\tau_L)+\widetilde{G}(\tau)$, where $\widetilde{G}$ is a known strictly increasing function such that $\widetilde{G}(\tau_L)=0$. Similar assumptions hold for $\delta(\cdot)$.

We now define the population objective function, for all functions $\xi(\cdot):(0,1)\rightarrow\mathbb{R}^p$, as:

$$Q\left(\xi(\cdot)\right) = \int_0^1 \left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau\right)\right]\right\|^2 d\tau. \tag{31}$$

[13]Note that the set of knots $\tau_1,...,\tau_L$ depends on L, but that we omit the dependence for conciseness. A more exact notation would be $\tau_{1L},...,\tau_{LL}$.

Moreover, we let $L = L_N$ depend on N, and we define the empirical objective function as:

$$Q_N(\xi(\cdot)) \quad = \quad \frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\widehat{\mathbb{E}}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2, \tag{32}$$

where $\widehat{\mathbb{E}}\left(Z_i\right) = \sum_{i=1}^N Z_i/N$.

In order to establish consistency of the REQR estimator we make the following assumption, where we let $L_N \to \infty$ as $N \to \infty$.

Assumption 6 *(consistency)*

(i) $Q(\xi(\cdot))$ is uniquely minimized at $\xi(\cdot) = \xi^0(\cdot)$ on $\mathcal{H}$.

(ii) The REQR estimator satisfies $\widehat{\xi}(\cdot) \in \mathcal{H}^{L_N}$ with probability approaching one, and:

$$Q_N(\widehat{\xi}(\cdot)) = \inf_{\xi(\cdot)\in\mathcal{H}^{L_N}} Q_N\left(\xi(\cdot)\right) + o_p(1).$$

(iii) Using the convention $\tau_0 \equiv 0$ and $\tau_{L_N+1} \equiv 1$, we have:

$$\sup_{\ell\in\{1,\ldots,L_N\}} L_N\left(\tau_{\ell+1}-\tau_\ell\right) \overset{N\to\infty}{\to} 1, \qquad \inf_{\ell\in\{1,\ldots,L_N\}} L_N\left(\tau_{\ell+1}-\tau_\ell\right) \overset{N\to\infty}{\to} 1.$$

(iv) There is a constant $M > 0$ such that $\|X_{it}\| \leq M$ with probability one.

Part (i) in Assumption 6 is an identification condition. In particular, this assumption is related to the completeness-type conditions discussed in Section 2.2.[14] Part (ii) requires that the REQR estimator asymptotically satisfies the conditions in Assumption 5.[15] We will impose (i), (iii), and (vi) in estimation, but not the inequality constraints (ii), (iv), and (v). Alternatively, note that, given choices for $\underline{c}$ and $\overline{c}$, these inequality constraints could be imposed in estimation, albeit at an increased computational cost—as the τ_ℓ-by-τ_ℓ nature of the M-step in the iterative algorithm would then be lost. Part (iii) allows for non-equidistant knots, but requires that they become equidistant in the limit. Finally, part (iv) requires covariates to have bounded support. Boundedness of the support is commonly assumed in the quantile regression literature. Note that the implied random variable η_i then

[14]However, note that Assumption 6 (ii) is *not* an implication of Proposition 1. The reason is that the identification theorem in Hu and Schennach (2008) relies on the full likelihood of the model, while (30) is a set of check-function based moment conditions.

[15]Note that the constants $(\overline{c},\underline{c})$ in $\mathcal{H}$ and $\mathcal{H}^{L_N}$ may be different. See the proof of Proposition 2 for detail.

also has bounded support.[16] Note also that the implied conditional densities of $Y_{it}|X_{it},\eta_i$ and $\eta_i|X_i$ are uniformly bounded from below and above.[17]

We now can state the consistency result.

Proposition 2 *(consistency) Let Assumption 6 hold. Then, as N and L_N tend to infinity so that L_N/N tends to zero, we have*

$$\sup_{\tau\in(\tau_1,\tau_{L_N})}\left\|\widehat{\xi}(\tau)-\xi^0(\tau)\right\|=o_p(1). \tag{33}$$

Proof. See Appendix B. ■

4.2 Parametric inference

As the REQR estimator is based on a method-of-moments approach, its large-N fixed-L asymptotic distribution is easily obtained by appealing to general results on GMM estimation with non-smooth moments (e.g., Newey and McFadden, 1994). This analysis will provide asymptotically valid inference provided the true parameters $\theta^0(\cdot)$ and $\delta^0(\cdot)$ are piecewise-linear functions with knots $\tau_1,...,\tau_L$. This is less general than the assumptions of Proposition 2, where the number of knots $L=L_N$ was allowed to increase with the sample size in order to approximate flexible population functions.

Here $\xi=(\xi_A',\xi_B')'$ is a $(pL)\times 1$ parameter vector. Stacking the pL moment restrictions of the approximating model given by (23)-(24) we obtain the following system:

$$\mathbb{E}\left[\Psi_i\left(\xi^0,\tau_\ell\right)\right] \;=\; 0, \quad \ell\in\{1,...,L\}. \tag{34}$$

Note that, as shown in the proof of Proposition 2, $\Psi_i\left(\xi,\tau_\ell\right)$ is smooth in its first argument, due to the presence of the integral and the fact that γ is bounded away from zero. Standard application of GMM then yields the following result.

Proposition 3 *(Asymptotic distribution) Under suitable conditions, and as N tends to infinity and L is kept fixed we have*

$$\sqrt{N}\left(\widehat{\xi}-\xi^0\right)\xrightarrow{d}\mathcal{N}\left(0,H_\xi^{-1}\Omega_\xi H_\xi^{-1}\right), \tag{35}$$

[16] Indeed, by (7) we have, for $\xi^0(\cdot)\in\mathcal{H}$: $|\eta_i|=\left|X_i'\delta^0(V_i)\right|\leq\|X_i\|\sup_{\tau\in(0,1)}\left\|\delta^0(\tau)\right\|<\infty$.

[17] For example we have: $\frac{\left|x'\left[\delta^0(\tau_2)-\delta^0(\tau_1)\right]\right|}{|\tau_2-\tau_1|}\leq\frac{\|x\|\left\|\delta^0(\tau_2)-\delta^0(\tau_1)\right\|}{|\tau_2-\tau_1|}\leq M\overline{c}$. Hence, by (21), $f_{\eta|X}(\eta|x)\geq\frac{1}{M\overline{c}}$.

where, for $\Psi_i(\xi,\cdot) = \left[\Psi_i(\xi,\tau_1)',...,\Psi_i(\xi,\tau_L)'\right]'$

$$H_\xi = \mathbb{E}\left[\frac{\partial \Psi_i(\xi^0,\cdot)}{\partial \xi}\frac{\partial \Psi_i(\xi^0,\cdot)}{\partial \xi'}\right], \quad \textit{and} \quad \Omega_\xi = \mathbb{E}\left[\Psi_i(\xi^0,\cdot)\,\Psi_i(\xi^0,\cdot)'\right].$$

An interesting question is to generalize Proposition 3 in order to allow L to tend to infinity with N. A special difficulty is caused by the fact that the model is subject to an ill-posed inverse problem. To see this, note that model (6)-(7) contains the following model as a special case

$$Y_{it} = \varepsilon_{it} + \eta_i, \tag{36}$$

where the only covariate is the constant term, where $\delta(\cdot) = 1$, and where we have denoted $\varepsilon_{it} = \beta(U_{it})$. Note that U_{it} is independent of η_i and U_{is}, $s \neq t$. Model (36) is a nonparametric deconvolution model with repeated measurements (Kotlarski, 1967). This model has been widely studied in statistics and econometrics (e.g., Horowitz and Markatou, 1996, Delaigle, Hall and Meister, 2008, Bonhomme and Robin, 2010). Hall and Lahiri (2008) show that in general quantiles of η_i and ε_{it} are not root-N estimable. Characterizing asymptotic rates of convergence and distributions in the more general model (6)-(7) is left to future work.

Finally, we note that extending the asymptotic results in this section to let L_N tend to infinity with N is important to provide guidance on the choice of number of knots in practice.

5 Implementation

In this section we describe our specific implementation of the REQR estimator based on piecewise-linear splines, and a simulation-based version of the fixed-point iterative algorithm introduced in Section 3.3.

5.1 Piecewise-linear splines

When using piecewise-linear splines as an approximating model, we obtain, for all $\ell = 1,...,L-1$:

$$\begin{aligned} \theta(\tau) &= \theta(\tau_\ell) + \frac{\tau-\tau_\ell}{\tau_{\ell+1}-\tau_\ell}\left[\theta(\tau_{\ell+1}) - \theta(\tau_\ell)\right], \quad \tau_\ell < \tau \leq \tau_{\ell+1}, \\ \delta(\tau) &= \delta(\tau_\ell) + \frac{\tau-\tau_\ell}{\tau_{\ell+1}-\tau_\ell}\left[\delta(\tau_{\ell+1}) - \delta(\tau_\ell)\right], \quad \tau_\ell < \tau \leq \tau_{\ell+1}. \end{aligned}$$

In this case, the implied approximating period-t density of outcomes and the implied approximating density of individual effects take closed-form expressions. Specifically, we have:

$$
\begin{aligned}
f\left(y_t \mid x_t, \eta; \xi_A\right) &= \frac{\tau_{\ell+1}-\tau_\ell}{w_t(\eta)'\left[\theta\left(\tau_{\ell+1}\right)-\theta\left(\tau_\ell\right)\right]} \quad \text{if } w_t(\eta)'\theta\left(\tau_\ell\right) < y_t \leq w_t(\eta)'\theta\left(\tau_{\ell+1}\right), \\
f\left(\eta \mid x; \xi_B\right) &= \frac{\tau_{\ell+1}-\tau_\ell}{x'\left[\delta\left(\tau_{\ell+1}\right)-\delta\left(\tau_\ell\right)\right]} \quad \text{if } x'\delta\left(\tau_\ell\right) < \eta \leq x'\delta\left(\tau_{\ell+1}\right),
\end{aligned}
$$

augmented with a specification in the extreme intervals $(0,\tau_1)$ and $(\tau_L,1)$.[18]

In order to model quantile functions in the intervals $(0,\tau_1)$ and $(\tau_L,1)$ one may assume, following Wei and Carroll (2009), that $\theta(\cdot)$ and $\delta(\cdot)$ are constant on these intervals, so the implied distribution functions have mass points at the two ends of the support. In Appendix A we outline a different, Laplace-based modelling of the extreme intervals, motivated by the desire to avoid that the support of the likelihood function depends on the parameter value. We use this method in the numerical and empirical exercises below.

5.2 Simulation-based algorithm

In practice we use a simulation-based approach in the "E"-step of the estimation algorithm. This allows us to replace the integrals in (28)-(29) by sums. Starting with initial parameter values $\widehat{\xi}^{(0)}$ we iterate the following two steps until convergence to a stationary distribution.

1. For all $i=1,...,N$, draw M values $\eta_i^{(1)}$,...,$\eta_i^{(M)}$ as follows, starting from an initial $\eta_i^{(0)}$:

    - Given $\eta_i^{(m-1)}$, draw $\widetilde{\eta}$ from $f\left(\eta \mid x_i; \widehat{\xi}_B^{(s)}\right)$.

      Set:
      $$r=\prod_{t=1}^{T}\frac{f\left(y_{it} \mid x_{it}, \widetilde{\eta}; \widehat{\xi}_A^{(s)}\right)}{f\left(y_{it} \mid x_{it}, \eta_i^{(m-1)}; \widehat{\xi}_A^{(s)}\right)}.$$

    - Set $\eta_i^{(m)}=\widetilde{\eta}$ with probability $\min(r,1)$.
    - Set $\eta_i^{(m)}=\eta_i^{(m-1)}$ with probability $1-\min(r,1)$.

2. Solve, for $\ell=1,...,L$:

$$
\begin{aligned}
\widehat{\theta}\left(\tau_\ell\right)^{(s+1)} &= \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{t=1}^{T}\rho_{\tau_\ell}\left(y_{it}-w_{it}\left(\eta_i^{(m)}\right)'\theta\right), \\
\widehat{\delta}\left(\tau_\ell\right)^{(s+1)} &= \underset{\delta}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{m=1}^{M}\rho_{\tau_\ell}\left(\eta_i^{(m)}-x_i'\delta\right).
\end{aligned}
$$

[18]Note that one could use a different spline family for implementation, although computing the implied likelihood functions would then require inverting the quantile functions numerically.

Step one—the "E" step—in this algorithm uses an independent Metropolis-Hastings method to draw from the posterior density $f\left(\eta \mid y_i, x_i; \widehat{\xi}^{(s)}\right)$, using the random-effects density $f\left(\eta \mid x_i; \widehat{\xi}_B^{(s)}\right)$ as proposal density. Note that draws from the latter distribution are easily obtained from (7) as $\widetilde{\eta} = x_i'\delta(\widetilde{v})$, where $\widetilde{v}$ is a standard uniform draw. Step 2—the "M" step—consists of $2L$ ordinary quantile regressions, where the simulated values of the individual effects are treated as covariates and dependent variables, respectively.

An advantage of Metropolis-Hastings over grid approximations and importance sampling weights is that the integral in the denominator of the posterior density of η is not needed. The output of this algorithm is a Markov chain. As M tends to infinity the sum converges to the true integral. The problem is then smooth (because of the integral with respect to η) and there should be no problem of non-convergence either.[19]

Note that this algorithm extends easily to models with multiple individual effects such as (9)-(10). Due to the triangular structure of the quantile models of the various components of η_i, the "E"-step in the above algorithm then simply consists in recursively drawing q different Markov chains, where q is the dimension of η_i.

6 Dynamic models

In this section we show how to extend the baseline REQR model (6)-(7) to dynamic settings. We consider three extensions in turn: autoregressive models, models with general predetermined regressors, and models with autocorrelated errors.

6.1 Autoregressive models

We start by considering a dynamic extension of model (3), where the conditional distribution of Y_{it} given past values $Y_i^{t-1} = (Y_{i,t-1}, ..., Y_{i1})$, strictly exogenous variables X_i, and individual effects η_i, is specified as follows:

$$Y_{it} = Q_Y\left(Y_{i,t-1}, X_{it}, \eta_i, U_{it}\right). \tag{37}$$

For simplicity here we focus on first-order autoregressive models. The extension to higher-order Markovian models is conceptually straightforward.

[19] In practice, the algorithm is based on a finite—relatively large—number of simulations. Assessing the impact of the choice of M on the statistical properties of the estimator is not done in this paper.

Identification. Nonparametric identification requires $T \geq 4$. Under Assumption 1, U_{it} is independent of X_{is} for all s and uniformly distributed, and independent of U_{is} for all $s \neq t$. So taking $T = 4$ we have:

$$\begin{aligned} f_{Y_2,Y_3,Y_4|Y_1,X}\left(y_2, y_3, y_4 \mid y_1, x\right) &= \int f_{Y_2|Y_1,\eta,X}\left(y_2 \mid y_1, \eta, x\right) f_{Y_3|Y_2,\eta,X}\left(y_3 \mid y_2, \eta, x\right) \\ &\quad \times f_{Y_4|Y_3,\eta,X}\left(y_4 \mid y_3, \eta, x\right) f_{\eta|Y_1,X}\left(\eta \mid y_1, x\right) d\eta, \end{aligned} \tag{38}$$

where we have used that Y_{i4} is conditionally independent of (Y_{i2}, Y_{i1}) given (Y_{i3}, X_i, η_i), and that Y_{i3} is conditionally independent of Y_{i1} given (Y_{i2}, X_i, η_i).

An extension of Hu and Schennach (2008)'s theorem, along the lines of Hu and Shum (2012), then shows nonparametric identification of all conditional densities $f_{Y_2|Y_1,\eta,X}$, $f_{Y_3|Y_2,\eta,X}$, $f_{Y_4|Y_3,\eta,X}$, and $f_{\eta|Y_1,X}$, in the autoregressive model, under suitable assumptions. A brief sketch of the identification argument is provided in Appendix C.[20]

Estimation. A dynamic counterpart of model (6) is given by:

$$Y_{it} = h\left(Y_{i,t-1}\right)' \alpha\left(U_{it}\right) + X_{it}' \beta\left(U_{it}\right) + \gamma\left(U_{it}\right) \eta_i, \quad t = 2, ..., T, \tag{39}$$

where

$$U_{it} \mid Y_i^{t-1}, X_i, \eta_i \sim \mathcal{U}(0, 1). \tag{40}$$

For example, when $h(y) = |y|$ model (39) is a panel data version of the CAViaR model of Engle and Manganelli (2004). Other choices will lead to panel counterparts of different dynamic quantile models (e.g., Gouriéroux and Jasiak, 2008).

Moreover, it is natural to let the distribution of η_i depend on initial conditions as well as strictly exogenous regressors; that is, extending (7):

$$\eta_i = \delta_2\left(V_i\right) Y_{i1} + X_i' \delta_1\left(V_i\right), \tag{41}$$

where

$$V_i \mid Y_{i1}, X_i \sim \mathcal{U}(0, 1). \tag{42}$$

[20] One issue is the choice of normalization. In the dynamic model (39), it follows from Hu and Shum (2012)'s analysis that one can use the same normalization (14) as in the static case, provided the averages across τ values of the coefficients of exogenous regressors and lagged outcome are identified based on:

$$\mathbb{E}\left[Y_{it} - Y_{i,t-1} \mid Y_i^{t-2}, X_i\right] = \mathbb{E}\left[h\left(Y_{i,t-1}\right) - h\left(Y_{i,t-2}\right) \mid Y_i^{t-2}, X_i\right]' \int_0^1 \alpha(\tau) d\tau + \left(X_{it} - X_{i,t-1}\right)' \int_0^1 \beta(\tau) d\tau.$$

The implied conditional quantile functions are, for $\tau \in (0,1)$:

$$\begin{aligned} Q_Y\left(Y_{i,t-1}, X_{it}, \eta_i, \tau\right) &= h\left(Y_{i,t-1}\right)' \alpha\left(\tau\right) + X_{it}'\beta\left(\tau\right) + \gamma\left(\tau\right)\eta_i, \\ Q_\eta\left(Y_{i1}, X_i, \tau\right) &= \delta_2\left(\tau\right) Y_{i1} + X_i'\delta_1\left(\tau\right). \end{aligned}$$

The estimation algorithm described in Section 3 applies to the autoregressive REQR model, provided the posterior density is now computed as:

$$f\left(\eta \mid y, x; \xi\right) = \frac{\prod_{t=2}^T f\left(y_t \mid y_{t-1}, x_t, \eta; \xi_A\right) f\left(\eta \mid y_1, x; \xi_B\right)}{\int \prod_{t=1}^T f\left(y_t \mid y_{t-1}, x_t, \widetilde{\eta}; \xi_A\right) f\left(\widetilde{\eta} \mid y_1, x; \xi_B\right) d\widetilde{\eta}}. \tag{43}$$

6.2 General predetermined regressors

We now consider model (39), but relax the assumption that X_{it} are strictly exogenous—as specified in equation (40)—and allow for general predetermined regressors. This leads to

$$U_{it} \mid Y_i^{t-1}, X_i^t, \eta_i \sim \mathcal{U}\left(0,1\right), \tag{44}$$

where the difference with (40) comes from the fact that current values of U_{it} may affect future values of covariates X_{is}, $s > t$.

The critical difference with the model with strictly exogenous regressors is in the nature of the posterior density of the individual effects. Indeed, let $W_{it} = (Y_{it}, X_{it}')'$, and let $W_i = (W_{i1}', ..., W_{iT}')'$. We have:

$$\begin{aligned} f\left(\eta \mid w; \xi\right) &= \frac{f\left(w_2, ..., w_T \mid w_1, \eta\right) f\left(\eta \mid w_1\right)}{\displaystyle\int f\left(w_2, ..., w_T \mid w_1, \eta\right) f\left(\eta \mid w_1\right) d\eta} \\ &= \frac{f\left(\eta \mid w_1; \xi_B\right) \prod_{t=2}^T f\left(y_t \mid y_{t-1}, x_t, \eta; \xi_A\right) f\left(x_t \mid w^{t-1}, \eta; \xi_C\right)}{\displaystyle\int f\left(\widetilde{\eta} \mid w_1; \xi_B\right) \prod_{t=2}^T f\left(y_t \mid y_{t-1}, x_t, \widetilde{\eta}; \xi_A\right) f\left(x_t \mid w^{t-1}, \widetilde{\eta}; \xi_C\right) d\widetilde{\eta}}, \end{aligned}$$

where now $\xi = (\xi_A', \xi_B', \xi_C')'$ includes additional parameters that correspond to the model of the *feedback process* from past values of Y_{it} and X_{it} to future values of X_{is}, for $s > t$.

Under predeterminedness, the structural quantile regression model only specifies the partial likelihood: $\prod_{t=2}^T f\left(y_t \mid y_{t-1}, x_t, \eta; \xi_A\right)$. However, the posterior density of the individual effects also depends on the feedback process $f\left(x_t \mid w^{t-1}, \eta; \xi_C\right)$, in addition to the random-effects density. In line with our approach, we also specify the feedback process as a quantile regression model. Specifically, letting X_{it}^k, $k = 1, ..., K$, denote the various components of

X_{it}, we specify the following triangular, recursive system:

$$\left\{ \begin{array}{lcl} X_{it}^1 & = & \mu_{12}\left(E_{i1}\right)Y_{i,t-1} + \mu_{11}\left(E_{i1}\right)'X_{it} + \mu_{10}\left(E_{i1}\right)\eta_i, \\ \cdots & \cdots & \cdots \\ X_{it}^K & = & \mu_{K3}\left(E_{iK}\right)'\left(X_{it}^1, ..., X_{it}^{K-1}\right)' + \mu_{K2}\left(E_{iK}\right)Y_{i,t-1} + \mu_{K1}\left(E_{iK}\right)'X_{it} + \mu_{K0}\left(E_{iK}\right)\eta_i, \end{array} \right. \tag{45}$$

where $E_{i1}, ..., E_{iK}$ follow independent standard uniform distributions, independent of all other random variables in the model, and where ξ_C stacks all μ parameters together.

Under Markovian assumptions, nonparametric identification may be established using similar arguments as in the strictly exogenous and autoregressive cases. Moreover, note that the triangular model (45) could depend on an additional vector of individual effects different from η_i, with no change to the general approach.

The model with predetermined regressors has thus three layers of quantile regressions: the outcome model (39), the model of the feedback process (45), and the model of individual effects, which now depends on the initial conditions of outcomes and covariates:

$$\eta_i = \delta_2\left(V_i\right)Y_{i1} + X_{i1}'\delta_1\left(V_i\right). \tag{46}$$

The estimation algorithm is similar to the one described in Section 3, with minor differences in the "E"- and "M"-steps.

6.3 Autocorrelated disturbances

As a third dynamic panel data extension, we next study a model with autocorrelated errors. Let us consider the baseline REQR model (6)-(7) with strictly exogenous regressors. We replace the independence assumption 1 *(iii)* by the following:

Assumption 7 *(autocorrelated errors)* $(U_1, ..., U_T)$ *is distributed as a copula* $C(u_1, ..., u_T)$ *given* X_i *and* η_i*.*

Nonparametric identification of the model (including the copula) can be shown under Markovian assumptions, as in the autoregressive model. For estimation we assume that the copula depends on a finite-dimensional parameter ϕ, which we estimate along with all quantile parameters.

The iterative estimation algorithm of Section 3 is easily modified by adding an extra "M" step where ϕ is updated as follows:

$$\widehat{\phi}^{(s+1)} = \underset{\phi}{\operatorname{argmax}} \sum_{i=1}^{N} \int \ln \left[c \left(F \left(y_{i1} | x_{i1}, \eta; \widehat{\xi}_A^{(s+1)} \right), ..., F \left(y_{iT} | x_{iT}, \eta; \widehat{\xi}_A^{(s+1)} \right); \phi \right) \right] \times f \left(\eta | y_i, x_i; \widehat{\xi}^{(s)}, \widehat{\phi}^{(s)} \right) d\eta, \tag{47}$$

where $c(u_1, ..., u_T) \equiv \partial^T C(u_1, ..., u_T)/\partial u_1 ... \partial u_T$ is the copula density, and where, for any y_t such that $w_t(\eta)' \theta(\tau_\ell) < y_t \leq w_t(\eta)' \theta(\tau_{\ell+1})$:

$$F(y_t | x_t, \eta; \xi_A) = \tau_\ell + (\tau_{\ell+1} - \tau_\ell) \frac{y_t - w_t(\eta)' \theta(\tau_\ell)}{w_t(\eta)' [\theta(\tau_{\ell+1}) - \theta(\tau_\ell)]},$$

augmented with a specification outside the interval $\left(w_t(\eta)' \theta(\tau_1), w_t(\eta)' \theta(\tau_L) \right)$.

The posterior density is then given by:

$$f(\eta | y, x; \xi, \phi) = \frac{\prod_{t=1}^T f(y_t \mid x_t, \eta; \xi_A)\, c\left[F(y_1 | x_1, \eta; \xi_A), ..., F(y_T | x_T, \eta; \xi_A); \phi \right] f(\eta \mid x; \xi_B)}{\int \prod_{t=1}^T f(y_t \mid x_t, \widetilde{\eta}; \xi_A)\, c\left[F(y_1 | x_1, \widetilde{\eta}; \xi_A), ..., F(y_T | x_T, \widetilde{\eta}; \xi_A); \phi \right] f(\widetilde{\eta} \mid x; \xi_B)\, d\widetilde{\eta}}.$$

Note that (47) boils down to a weighted least-squares regression when using a bivariate Gaussian copula.[21] Note also that allowing for autocorrelated disturbances in autoregressive models such as (39)—that is, for ARMA-type quantile regression models—is not straightforward.

7 Numerical and empirical results

In this section, we apply the REQR estimator to simulated data and to an empirical question and dataset.

7.1 Monte Carlo exercise

The simulated model is as follows:

$$Y_{it} = \beta_0(U_{it}) + \beta_1(U_{it}) X_{1it} + \beta_2(U_{it}) X_{2it} + \gamma(U_{it}) \eta_i,$$

[21] Specifically, in the "M"-step of the estimation algorithm, $\widehat{\phi}^{(s+1)}$ may then be recovered using a pooled regression of:

$$\Phi^{-1} \left[F \left(y_{it} | x_{it}, \eta_i^{(m)}; \widehat{\xi}_A^{(s)} \right) \right] \quad \text{on} \quad \Phi^{-1} \left[F \left(y_{i,t-1} | x_{i,t-1}, \eta_i^{(m)}; \widehat{\xi}_A^{(s)} \right) \right],$$

where $\Phi(\cdot)$ denotes the standard normal cdf. Note that, for this regression to be formally identical to the solution of (47), one would need to also impose that the error in the regression has variance $1 - \phi^2$.

Table 1: Monte Carlo results

τ	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{4}{12}$	$\frac{5}{12}$	$\frac{6}{12}$	$\frac{7}{12}$	$\frac{8}{12}$	$\frac{9}{12}$	$\frac{10}{12}$	$\frac{11}{12}$
	Quantile parameters: outcomes										
	Population values										
Cst	-.719	-.483	-.330	-.208	-.101	.000	.101	.208	.330	.483	.719
X_1	.281	.517	.670	.792	.899	1.00	1.10	1.21	1.33	1.48	1.72
X_2	.281	.517	.670	.792	.899	1.00	1.10	1.21	1.33	1.48	1.72
η	.640	.759	.835	.896	.950	1.00	1.05	1.10	1.16	1.24	1.36
	Monte Carlo means										
Cst	-.993	-.592	-.398	-.253	-.126	-.009	.120	.264	.412	.604	.970
X_1	.300	.507	.673	.795	.902	1.01	1.12	1.22	1.36	1.52	1.70
X_2	.294	.497	.680	.793	.893	1.00	1.11	1.22	1.35	1.52	1.71
η	.727	.792	.857	.912	.959	1.00	1.04	1.08	1.13	1.21	1.29
	Monte Carlo standard deviations										
Cst	.216	.171	.154	.135	.117	.106	.098	.104	.143	.211	.319
X_1	.081	.100	.086	.078	.075	.068	.075	.089	.099	.103	.086
X_2	.075	.106	.075	.083	.086	.077	.073	.069	.084	.090	.088
η	.073	.054	.047	.040	.036	.031	.030	.036	.045	.061	.091
	Quantile parameters: individual effects										
	Population values										
Cst	2.14	2.26	2.34	2.40	2.44	2.50	2.55	2.60	2.66	2.74	2.86
X_1	.140	.259	.335	.396	.450	.500	.551	.604	.665	.741	.860
X_2	.140	.259	.335	.396	.450	.500	.551	.604	.665	.741	.860
	Monte Carlo means										
Cst	1.94	2.21	2.32	2.39	2.44	2.49	2.53	2.59	2.66	2.78	3.06
X_1	.162	.259	.336	.394	.447	.495	.538	.586	.644	.724	.837
X_2	.137	.240	.324	.394	.453	.508	.562	.617	.682	.767	.877
	Monte Carlo standard deviations										
Cst	.240	.155	.137	.134	.131	.127	.123	.124	.141	.178	.261
X_1	.160	.123	.112	.109	.109	.111	.116	.112	.121	.140	.189
X_2	.200	.151	.123	.110	.109	.113	.115	.116	.123	.143	.195

Notes: $N = 1000$, $T = 3$, 100 *simulations (*100 *iterations of the sequential algorithm,* 50 *draws per individual in each simulation.)*

where X_{1it} and X_{2it} follow independent χ_1^2 distributions, and where U_{it} are i.i.d., uniform on the unit interval. Individual effects are generated as

$$\eta_i = \delta_0(V_i) + \delta_1(V_i)\overline{X}_{1i} + \delta_2(V_i)\overline{X}_{2i},$$

where V_i is i.i.d. uniform on $(0,1)$, independent of everything else, and where $\overline{X}_{1i}$ and $\overline{X}_{2i}$ denote individual averages. Lastly, $\beta_j(\tau)$ and $\delta_j(\tau)$ are defined on a set of $L = 11$ knots; see top row of Table 1. Here, as in the empirical analysis below, we use linear splines to construct the approximating model. We also use the Laplace modelling described in Appendix A for the extreme intervals. Lastly, we use the correct number of knots ($L = 11$).

The top panel in Table 1 shows the estimates of $\beta_0(\tau_\ell)$, $\beta_1(\tau_\ell)$, $\beta_2(\tau_\ell)$, and $\gamma(\tau_\ell)$, across 100 simulated datasets with $N = 1000$ and $T = 3$. We report the population values of the parameters, and means and standard deviations across simulations. The results show moderate finite-sample biases, and relatively precise estimates, even at the extreme knots. We also observe larger biases for $\gamma(\cdot)$ and the constant coefficient. The bottom panel in Table 1 shows the estimates of $\delta_0(\tau_\ell)$, $\delta_1(\tau_\ell)$, and $\delta_2(\tau_\ell)$. In this case we observe somewhat larger standard errors in the tails. Nevertheless, biases throughout the distribution are moderate. Together, these preliminary results suggest reasonable finite-sample performance of the REQR estimator.

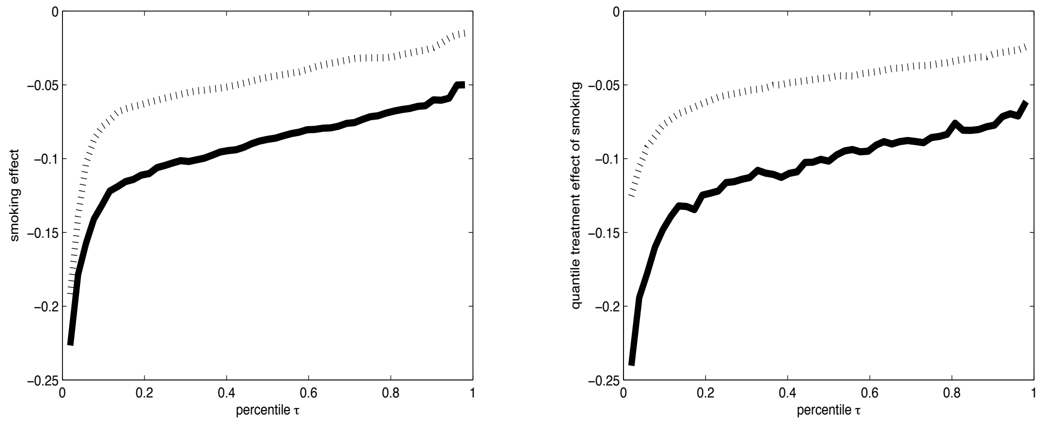
7.2 Empirical application

We revisit the effect of maternal inputs of children's birth outcomes. Specifically, we study the effect of smoking during pregnancy on children's birthweights. Abrevaya (2006) uses a mother fixed-effects approach to address endogeneity of smoking. Here we use quantile regression with mother-specific effects to allow for both unobserved heterogeneity and non-linearities in the relationship between smoking and weight at birth.

We use a balanced subsample from the US natality data used in Abrevaya (2006), which comprises 12360 women with 3 children each. Our outcome is the log-birthweight. The main covariate is a binary smoking indicator. Age of the mother and gender of the child are used as additional controls.

An OLS regression yields a negative point estimate of the smoking coefficient: $-.095$. The fixed-effects estimate is also negative, but it is twice as small: $-.050$ (significant). This suggests a negative endogeneity bias in OLS, and is consistent with Abrevaya (2006). The solid line on the left graph of Figure 1 shows the smoking coefficient estimated from pooled quantile regressions, on a fine grid of τ values. According to these estimates, the effect of smoking is more negative at lower quantiles of birthweights. The dashed line on the same graph shows the REQR estimate of the smoking effect.[22] We see that the smoking effect becomes less negative when correcting for time-invariant endogeneity through the introduction of mother-specific fixed-effects. At the same time, the effect is still sizable, and it remains increasing along the distribution.

Figure 1: Quantile effects of smoking during pregnancy on log-birthweight



Note: Data from Abrevaya (2006). Left graph: solid line is the pooled quantile regression smoking coefficient; dashed line is the panel quantile regression (REQR) smoking coefficient. Right graph: solid line is the raw quantile treatment effect of smoking; dashed line is the quantile treatment effect estimate based on REQR.

As a last exercise, on the right graph of Figure 1 we compute the *quantile treatment effect* of smoking as the difference in log-birthweights between a sample of smoking women, and a sample of non-smoking women, keeping all other characteristics—observed (X_i) and unobserved (η_i)—constant. This calculation illustrates the usefulness of specifying and estimating a complete semiparametric model of the joint distribution of outcomes and unobservables, in order to compute counterfactual distributions that take into account the presence of unobserved heterogeneity. On the graph, the solid line shows the empirical difference between unconditional quantiles, while the dashed line shows the quantile treatment effect that accounts for both observables and unobservables. The results are broadly similar to the ones

[22] REQR estimates are computed using $L = 21$ knots. The simulated EM-type algorithm has 100 iterations, with 50 Metropolis-hastings draws within each iteration.

reported on the left graph of Figure 1. An interesting finding is that in this case the endogeneity bias—i.e., the difference between the dashed and solid lines—decreases as one moves from lower to higher quantiles of birthweight.

8 Conclusion

Random-effects quantile regression (REQR) provides a flexible approach to model nonlinear panel data models. In our approach, quantile regression is used as a versatile tool to model the dependence between individual effects and exogenous regressors or initial conditions, and to model feedback processes in models with predetermined covariates. The empirical application illustrates the benefits of having a flexible approach to allow for heterogeneity and nonlinearity within the same model in a panel data context.

The analysis of the asymptotic properties of the REQR estimator requires an approximation argument. However, while our consistency proof allows the quality of the approximation to increase with the sample size, at this stage in our characterization of the asymptotic distribution we keep the number of knots L fixed as the number of observations N increases. Assessing the asymptotic behavior of the quantile estimates as both L and N tend to infinity is an important task for future work.

Lastly, note that our quantile-based modelling of the distribution of individual effects could be of interest in other models as well. For example, one could consider semiparametric likelihood panel data models, where the conditional likelihood of the outcome Y_i given X_i and η_i depends on a finite-dimensional parameter vector α, and the conditional distribution of η_i given X_i is left unrestricted. The approach of this paper is easily adapted to this case, and delivers a semiparametric likelihood of the form:

$$f(y_i|x_i;\alpha,\delta(\cdot)) = \int f(y_i|x_i,\eta_i;\alpha) f(\eta_i|x_i;\delta(\cdot)) d\eta_i,$$

where $\delta(\cdot)$ is a process of quantile coefficients.

As another example, our framework naturally extends to models with time-varying unobservables, such as:

$$\begin{aligned} Y_{it} &= Q_Y\left(X_{it}, \eta_{it}, U_{it}\right), \\ \eta_{it} &= Q_\eta\left(\eta_{i,t-1}, V_{it}\right), \end{aligned}$$

where U_{it} and V_{it} are i.i.d. and uniformly distributed. It seems worth assessing the usefulness of our approach in these other contexts.

References

[1] Abrevaya, J. (2006): "Estimating the Effect of Smoking on Birth Outcomes Using a Matched Panel Data Approach," *Journal of Applied Econometrics*, vol. 21(4), 489–519.

[2] Abrevaya, J., and C. M. Dahl (2008): "The Effects of Birth Inputs on Birthweight," *Journal of Business & Economic Statistics*, 26, 379–397.

[3] Ai, C. and Chen, X. (2003): "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, 71, 1795–1843.

[4] Andrews, D. (2011): "Examples of L^2-Complete and Boundedly-Complete Distributions," unpublished manuscript.

[5] Arcidiacono, P. and J. B. Jones (2003): "Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm," *Econometrica*, 71(3), 933–946.

[6] Arcidiacono, P. and R. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 7, 1823–1868.

[7] Arellano, M. and S. Bonhomme (2011): "Nonlinear Panel Data Analysis," *Annual Review of Economics*, 3, 2011, 395-424.

[8] Arellano, M. and S. Bonhomme (2012): "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," *Review of Economic Studies*, 79, 987–1020.

[9] Belloni, A., Chernozhukov, V., and I. Fernández-Val (2011): "Conditional Quantile Processes based on Series or Many Regressors," unpublished manuscript.

[10] Bonhomme, S. and J.-M. Robin (2009): "Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990-2000," *Review of Economic Studies*, 76, 63–92.

[11] Bonhomme, S. and J.-M. Robin (2010): "Generalized Nonparametric Deconvolution with an Application to Earnings Dynamics," *Review of Economic Studies*, 77, 491- 533.

[12] Bonhomme, S. (2012): "Functional Differencing," *Econometrica*, 80, 1337–1385.

[13] Cai, Z. and X. Xu (2008): "Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models," *Journal of the American Statistical Association*, 103, 1595–1608.

[14] Canay, I. A. (2011): “A Simple Approach to Quantile Regression for Panel Data,” *The Econometrics Journal*, 14 (3), 368–386.

[15] Canay, I. A., A. Santos, and A. Shaikh (2012): “On the Testability of Identification in Some Nonparametric Models with Endogeneity”, unpublished manuscript.

[16] Chamberlain, G. (1984): “Panel Data”, in Griliches, Z. and M. D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, Elsevier Science, Amsterdam.

[17] Chaudhuri, P. (1991): “Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation,” *Annals of Statistics*, 19, 760–777.

[18] Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2012): “Average and Quantile Effects in Nonseparable Panel Models,” *Econometrica*, forthcoming.

[19] Delaigle, A., P. Hall, and A. Meister (2008): “On Deconvolution with Repeated Measurements,” *Annals of Statistics*, 36, 665–685.

[20] Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society*, B, 39, 1–38.

[21] D’Haultfoeuille, X. (2011): “On the Completeness Condition for Nonparametric Instrumental Problems,” *Econometric Theory*, 27, 460–471

[22] Elashoff, M. and L. Ryan (2004): “An EM Algorithm for Estimating Equations,” *Journal of Computational and Graphical Statistics*, 13(1), 48–65.

[23] Engle, R. F. and S. Manganelli (2004): “CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles,” *Journal of Business & Economic Statistics*, 22, 367–381.

[24] Evdokimov, K. (2010): “Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity,” unpublished manuscript.

[25] Gallant, A. R. and D. W. Nychka (1987): “Semi-Nonparametric Maximum Likelihood Estimation”, *Econometrica* 55, 363–390

[26] Galvao, A. F. (2011): ”Quantile Regression for Dynamic Panel Data with Fixed Effects,” *Journal of Econometrics*, 164, 142–157.

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

[28] Geraci, M., and M. Bottai (2007): "Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution," *Biostatistics*, 8(1), 140–154.

[29] Gouriéroux, C. and J. Jasiak (2008): "Dynamic Quantile Models," *Journal of Econometrics*, 147, 198–205.

[30] Hall, P. and S. N. Lahiri (2008): "Estimation of Distributions, Moments and Quantiles in Deconvolution Problems," *Annals of Statistics*, 36, 2110-2134.

[31] He, X. (1997): "Quantile curves without crossing," *The American Statistician*, 51, 186–192.

[32] Honoré, B., and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629.

[33] Horowitz, J. L., and M. Markatou (1996): "Semiparametric Estimation of Regression Models for Panel Data," *Review of Economic Studies*, 63, 145–168.

[34] Hu, Y. and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195–216.

[35] Hu, Y. and J.-L. Shiu (2012): "Nonparametric Identification Using Instrumental Variables: Sufficient Conditions for Completeness," unpublished manuscript.

[36] Hu, Y. and M. Shum (2012): "Nonparametric Identification of Dynamic Models with Unobserved State Variables," *Journal of Econometrics*, 171, 32–44.

[37] Koenker, R. and G. J. Bassett (1978): "Regression quantiles," *Econometrica*, 46, 33-50.

[38] Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

[39] Kotlarski, I. (1967): "On Characterizing the Gamma and Normal Distribution," *Pacific Journal of Mathematics*, 20, 69–76.

[40] Lamarche, C. (2010): "Robust Penalized Quantile Regression for Panel Data," *Journal of Econometrics*, 157, 396–408.

[41] Li, T., and Q. Vuong (1998): "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139–165.

[42] McLachlan, G. J., and T. Krishnan (2008): *The EM Algorithm and Extensions*, New York, Wiley, Second Edition.

[43] Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4, Elsevier.

[44] Newey, W. and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

[45] Qu, Z. and J. Yoon (2011): "Nonparametric Estimation and Inference on Conditional Quantile Processes," unpublished manuscript.

[46] Rosen, A. (2012): "Set Identification via Quantile Restrictions in Short Panels," *Journal of Econometrics*, 166, 127–137.

[47] Wei, Y. and R. J. Carroll (2009): "Quantile Regression with Measurement Error," *Journal of the American Statistical Association*, 104, 1129–1143.

[48] Wu, C. F. J. (1983): "On the Convergence Properties of the EM Algorithm," *Annals of Statistics*, 11, 95–103.

APPENDIX

A Laplace modelling of the tails

For implementation, we use the following modelling for the splines in the extreme intervals:

$$
\begin{aligned}
\theta(\tau) &= \theta(\tau_1) + \frac{\ln(\tau/\tau_1)}{1-\tau_1}\iota_c, \quad \tau \leq \tau_1, \\
\theta(\tau) &= \theta(\tau_L) - \frac{\ln(1-(\tau-\tau_L)/(1-\tau_L))}{\tau_L}\iota_c, \quad \tau > \tau_L,
\end{aligned}
$$

where ι_c is a vector of zeros, with a one at the position of the constant term in $\theta(\tau)$. We adopted a similar specification for $\delta(\tau)$. Modelling the constant terms in $\theta(\tau)$ and $\delta(\tau)$ as we do avoids the inconvenient that the support of the likelihood function depends on the parameter value. Moreover, our specification boils down to the Laplace model of Geraci and Bottai (2007) when $L = 1$.

The implied approximating period-t outcome density is then

$$
\begin{aligned}
f(y_t \mid x_t, \eta; \xi_A) &= \sum_{\ell=1}^{L-1} \frac{\tau_{\ell+1}-\tau_\ell}{w_t(\eta)'[\theta(\tau_{\ell+1})-\theta(\tau_\ell)]}\mathbf{1}\left\{w_t(\eta)'\theta(\tau_\ell) < y_t \leq w_t(\eta)'\theta(\tau_{\ell+1})\right\} \\
&\quad + \tau_1(1-\tau_1)e^{(1-\tau_1)\left(y_t - w_t(\eta)'\theta(\tau_1)\right)}\mathbf{1}\left\{y_t \leq w_t(\eta)'\theta(\tau_1)\right\} \\
&\quad + \tau_L(1-\tau_L)e^{-\tau_L\left(y_t - w_t(\eta)'\theta(\tau_L)\right)}\mathbf{1}\left\{y_t > w_t(\eta)'\theta(\tau_L)\right\}.
\end{aligned}
$$

Similarly, the approximating density of individual effects is

$$
\begin{aligned}
f(\eta \mid x; \xi_B) &= \sum_{\ell=1}^{L-1} \frac{\tau_{\ell+1}-\tau_\ell}{x'[\delta(\tau_{\ell+1})-\delta(\tau_\ell)]}\mathbf{1}\left\{x'\delta(\tau_\ell) < \eta \leq x'\delta(\tau_{\ell+1})\right\} \\
&\quad + \tau_1(1-\tau_1)e^{(1-\tau_1)(\eta - x'\delta(\tau_1))}\mathbf{1}\left\{\eta \leq x'\delta(\tau_1)\right\} \\
&\quad + \tau_L(1-\tau_L)e^{-\tau_L(\eta - x'\delta(\tau_L))}\mathbf{1}\left\{\eta > x'\delta(\tau_L)\right\}.
\end{aligned}
$$

B Proof of Proposition 2

We will denote:

$$
\|\xi(\cdot)\|_\infty = \sup_{\tau\in(0,1)} \|\xi(\tau)\|, \quad \text{for all } \xi(\cdot) \in \mathcal{H}.
$$

We start with two preliminary lemmas.

Lemma B1 *Let Assumption 6 hold. Then there exist three constants $0 < \nu < \frac{1}{2}$, $C_1 > 0$, and $C_2 > 0$, such that, with probability one:*

$$
\|\Psi_i(\xi_2(\cdot),\tau_2) - \Psi_i(\xi_1(\cdot),\tau_1)\| \leq C_1|\tau_2-\tau_1| + C_2\|\xi_2(\cdot)-\xi_1(\cdot)\|_\infty^\nu,
$$

for all $(\tau_1,\tau_2) \in (0,1)^2$ and all $(\xi_1(\cdot),\xi_2(\cdot)) \in \mathcal{H}\times\mathcal{H}$.

Proof.

We start by noting that:

$$
\begin{aligned}
\|\Psi_i(\xi_2(\cdot),\tau_2) - \Psi_i(\xi_1(\cdot),\tau_1)\| &\leq \|\Psi_i(\xi_2(\cdot),\tau_2) - \Psi_i(\xi_2(\cdot),\tau_1)\| \\
&\quad + \|\Psi_i(\xi_2(\cdot),\tau_1) - \Psi_i(\xi_1(\cdot),\tau_1)\|, \qquad \text{(B1)}
\end{aligned}
$$

and we will bound the two terms in (B1) in turn.

Starting with the first term we have, denoting $f_i(\eta;\xi(\cdot)) \equiv f(\eta \mid Y_i, X_i; \xi(\cdot))$ for conciseness:

$$\begin{aligned}
\left\|\Psi_i(\xi_2(\cdot),\tau_2)-\Psi_i(\xi_2(\cdot),\tau_1)\right\| \quad &\leq \quad \Big\| \int \sum_{t=1}^{T} W_{it}(\eta) \Big[\psi_{\tau_2}\left(Y_{it}-W_{it}(\eta)'\theta_2(\tau_2)\right) \\
&\qquad -\psi_{\tau_1}\left(Y_{it}-W_{it}(\eta)'\theta_2(\tau_1)\right)\Big] f_i(\eta;\xi_2(\cdot))\,d\eta \Big\| \\
&\quad +\Big\| \int X_i \Big[\psi_{\tau_2}\left(\eta-X_i'\delta_2(\tau_2)\right) \\
&\qquad -\psi_{\tau_1}\left(\eta-X_i'\delta_2(\tau_1)\right)\Big] f_i(\eta;\xi_2(\cdot))\,d\eta \Big\|.
\end{aligned}$$

We have:

$$\begin{aligned}
I \quad &\equiv \quad \Big\| \int \sum_{t=1}^{T} W_{it}(\eta) \Big[\psi_{\tau_2}\left(Y_{it}-W_{it}(\eta)'\theta_2(\tau_2)\right) \\
&\qquad -\psi_{\tau_1}\left(Y_{it}-W_{it}(\eta)'\theta_2(\tau_1)\right)\Big] f_i(\eta;\xi_2(\cdot))\,d\eta \Big\| \\
&\leq \quad \Big\| \int \sum_{t=1}^{T} W_{it}(\eta) f_i(\eta;\xi_2(\cdot))\,d\eta \Big\| \left|\tau_2-\tau_1\right| \\
&\quad +\Big\| \int \sum_{t=1}^{T} W_{it}(\eta) \Big[\mathbf{1}\left\{Y_{it}\leq X_{it}'\beta_2(\tau_2)+\gamma(\tau_2)\eta\right\} \\
&\qquad -\mathbf{1}\left\{Y_{it}\leq X_{it}'\beta_2(\tau_1)+\gamma(\tau_1)\eta\right\}\Big] f_i(\eta;\xi_2(\cdot))\,d\eta \Big\|.
\end{aligned}$$

Now, the posterior density $f_i(\eta;\xi_2(\cdot))$ is bounded from above. Indeed, because $\xi_2(\cdot)\in\mathcal{H}$, $f_{\eta|X;\delta_2(\cdot)}(\eta|x)\leq\frac{1}{\underline{c}}$, and similarly $f_{Y_t|\eta,X_t;\theta_2(\cdot)}(y_t|\eta,x_t)\leq\frac{1}{\underline{c}}$ for all t. Moreover, as $f_{Y_t|\eta,X_t;\theta_2(\cdot)}(y_t|\eta,x_t)\geq\frac{1}{M\overline{c}}$ we have:

$$\begin{aligned}
\int\prod_{t=1}^{T} f_{Y_t|\eta,X_t;\theta_2(\cdot)}(y_t|\eta,x_t) f_{\eta|X;\delta_2(\cdot)}(\eta|x)d\eta \quad &\geq \quad \left(\frac{1}{M\overline{c}}\right)^T \int f_{\eta|X;\delta_2(\cdot)}(\eta|x)d\eta \\
&= \quad \left(\frac{1}{M\overline{c}}\right)^T.
\end{aligned}$$

It thus follows from (15) that:

$$f_i(\eta;\xi_2(\cdot))\leq\frac{(M\overline{c})^T}{\underline{c}^{T+1}}<\infty. \tag{B2}$$

Now, as the support of $W_{it}(\eta)$ is uniformly bounded it follows that, for some constants $\widetilde{C}_1$ and $\widetilde{C}_2$ we have:

$$\begin{aligned}
I \quad &\leq \quad \widetilde{C}_1\left|\tau_2-\tau_1\right|+\widetilde{C}_2\int\Big|\mathbf{1}\left\{Y_{it}\leq X_{it}'\beta_2(\tau_2)+\gamma_2(\tau_2)\eta\right\}-\mathbf{1}\left\{Y_{it}\leq X_{it}'\beta_2(\tau_1)+\gamma_2(\tau_1)\eta\right\}\Big|d\eta \\
&= \quad \widetilde{C}_1\left|\tau_2-\tau_1\right|+\widetilde{C}_2\int\Big|\mathbf{1}\left\{\eta\geq\frac{Y_{it}-X_{it}'\beta_2(\tau_2)}{\gamma_2(\tau_2)}\right\}-\mathbf{1}\left\{\eta\geq\frac{Y_{it}-X_{it}'\beta_2(\tau_1)}{\gamma_2(\tau_1)}\right\}\Big|d\eta \\
&\leq \quad \widetilde{C}_1\left|\tau_2-\tau_1\right|+\widetilde{C}_2\Big|\frac{Y_{it}-X_{it}'\beta_2(\tau_2)}{\gamma_2(\tau_2)}-\frac{Y_{it}-X_{it}'\beta_2(\tau_1)}{\gamma_2(\tau_1)}\Big| \\
&\leq \quad \widetilde{C}_1\left|\tau_2-\tau_1\right|+\widetilde{C}_2\frac{\Big|\gamma_2(\tau_1)\left(Y_{it}-X_{it}'\beta_2(\tau_2)\right)-\gamma_2(\tau_2)\left(Y_{it}-X_{it}'\beta_2(\tau_1)\right)\Big|}{\underline{c}^2} \\
&\leq \quad \widetilde{\widetilde{C}}_1\left|\tau_2-\tau_1\right|,
\end{aligned}$$

where we have used that $\beta_2(\cdot)$ and $\gamma_2(\cdot)$ are Lipschitz— as $\xi_2(\cdot) \in \mathcal{H}$ —, that $\gamma_2(\cdot)$ is bounded from below, and that X_{it} and also Y_{it} have bounded support.[23]

Using similar arguments, it is easily verified that:

$$\begin{aligned} II &\equiv \Big\| \int X_i \Big[\psi_{\tau_2} \left(\eta - X_i' \delta_2 (\tau_2) \right) \\ &\qquad - \psi_{\tau_1} \left(\eta - X_i' \delta_2 (\tau_1) \right) \Big] f_i \left(\eta ; \xi_2(\cdot) \right) d\eta \Big\| \\ &\leq \widetilde{\widetilde{C}}_1 \left| \tau_2 - \tau_1 \right| , \end{aligned}$$

possibly with a different constant $\widetilde{\widetilde{C}}_1$.

It thus follows that, for some constant C_1:

$$\left\| \Psi_i \left(\xi_2(\cdot), \tau_2 \right) - \Psi_i \left(\xi_2(\cdot), \tau_1 \right) \right\| \leq C_1 \left| \tau_2 - \tau_1 \right| .$$

We next turn to the second term in (B1). We have:

$$\begin{aligned} \left\| \Psi_i \left(\xi_2(\cdot), \tau_1 \right) - \Psi_i \left(\xi_1(\cdot), \tau_1 \right) \right\| &\leq \Big\| \int \sum_{t=1}^T W_{it} (\eta) \, \psi_{\tau_1} \left(Y_{it} - W_{it} (\eta)' \theta_2 (\tau_1) \right) f_i \left(\eta ; \xi_2(\cdot) \right) d\eta \\ &\quad - \int \sum_{t=1}^T W_{it} (\eta) \, \psi_{\tau_1} \left(Y_{it} - W_{it} (\eta)' \theta_1 (\tau_1) \right) f_i \left(\eta ; \xi_1(\cdot) \right) d\eta \Big\| \\ &\quad + \Big\| \int X_i \psi_{\tau_1} \left(\eta - X_i' \delta_2 (\tau_1) \right) f_i \left(\eta ; \xi_2(\cdot) \right) d\eta \\ &\quad - \int X_i \psi_{\tau_1} \left(\eta - X_i' \delta_1 (\tau_1) \right) f_i \left(\eta ; \xi_1(\cdot) \right) d\eta \Big\| . \end{aligned}$$

As before we define:

$$\begin{aligned} I &\equiv \Big\| \int \sum_{t=1}^T W_{it} (\eta) \, \psi_{\tau_1} \left(Y_{it} - W_{it} (\eta)' \theta_2 (\tau_1) \right) f_i \left(\eta ; \xi_2(\cdot) \right) d\eta \\ &\quad - \int \sum_{t=1}^T W_{it} (\eta) \, \psi_{\tau_1} \left(Y_{it} - W_{it} (\eta)' \theta_1 (\tau_1) \right) f_i \left(\eta ; \xi_1(\cdot) \right) d\eta \Big\| \\ &\leq \Big\| \int \sum_{t=1}^T W_{it} (\eta) \Big[\psi_{\tau_1} \left(Y_{it} - W_{it} (\eta)' \theta_2 (\tau_1) \right) - \psi_{\tau_1} \left(Y_{it} - W_{it} (\eta)' \theta_1 (\tau_1) \right) \Big] f_i \left(\eta ; \xi_2(\cdot) \right) d\eta \Big\| \\ &\quad + \Big\| \int \sum_{t=1}^T W_{it} (\eta) \, \psi_{\tau_1} \left(Y_{it} - W_{it} (\eta)' \theta_1 (\tau_1) \right) \Big[f_i \left(\eta ; \xi_2(\cdot) \right) - f_i \left(\eta ; \xi_1(\cdot) \right) \Big] d\eta \Big\| . \end{aligned}$$

Now, using that $W_{it}(\eta)$ has bounded support, using that by (B2) $f_i(\eta; \xi_2(\cdot))$ is bounded from above, and using similar derivations as before, we obtain, for some constant $\widetilde{C}_2$:

$$\begin{aligned} I_1 &\equiv \Big\| \int \sum_{t=1}^T W_{it} (\eta) \Big[\psi_{\tau_1} \left(Y_{it} - W_{it} (\eta)' \theta_2 (\tau_1) \right) - \psi_{\tau_1} \left(Y_{it} - W_{it} (\eta)' \theta_1 (\tau_1) \right) \Big] f_i \left(\eta ; \xi_2(\cdot) \right) d\eta \Big\| \\ &\leq \widetilde{C}_2 \Big| \frac{Y_{it} - X_{it}' \beta_2 (\tau_1)}{\gamma_2(\tau_1)} - \frac{Y_{it} - X_{it}' \beta_1 (\tau_1)}{\gamma_1(\tau_1)} \Big| \\ &\leq \widetilde{C}_2 \frac{\Big| \gamma_1(\tau_1) \left(Y_{it} - X_{it}' \beta_2 (\tau_1) \right) - \gamma_2(\tau_1) \left(Y_{it} - X_{it}' \beta_1 (\tau_1) \right) \Big|}{\underline{c}^2} \\ &\leq \widetilde{\widetilde{C}}_2 \left\| \xi_2(\cdot) - \xi_1(\cdot) \right\|_\infty , \end{aligned}$$

[23]Indeed, by (6) we have: $|Y_{it}| \leq \|X_{it}\| \, \|\beta_2(\cdot)\|_\infty + |\eta_i| \, \|\gamma_2(\cdot)\|_\infty < \infty$.

where we have used that X_{it} and Y_{it} have bounded supports.

We then have:

$$\begin{aligned} I_2 &\equiv \left\| \int \sum_{t=1}^{T} W_{it}(\eta)\, \psi_{\tau_1}\left(Y_{it} - W_{it}(\eta)'\theta_1(\tau_1)\right) \left[f_i(\eta;\xi_2(\cdot)) - f_i(\eta;\xi_1(\cdot)) \right] d\eta \right\| \\ &\leq \widetilde{C}_2 \int |f_i(\eta;\xi_2(\cdot)) - f_i(\eta;\xi_1(\cdot))|\, d\eta, \end{aligned}$$

for a possibly different constant $\widetilde{C}_2 > 0$.

To show that $I_2 \leq \widetilde{\widetilde{C}}_2 \|\xi_2(\cdot) - \xi_1(\cdot)\|_\infty^\nu$, it suffices to bound $|f_i(\eta;\xi_2(\cdot)) - f_i(\eta;\xi_1(\cdot))|$. In turn, it is enough to bound the difference between the random-effects densities of individual effects associated with $\xi_2(\cdot)$ and $\xi_1(\cdot)$, as well as the differences between the period-specific likelihood functions of outcomes. Here we detail the argument for the former, the latter being similar.

By (21) we have:

$$f_{\eta|X;\delta_2(\cdot)}(\eta|x) - f_{\eta|X;\delta_1(\cdot)}(\eta|x) = \frac{1}{x'\nabla\delta_2\left(F[\eta|x;\delta_2(\cdot)]\right)} - \frac{1}{x'\nabla\delta_1\left(F[\eta|x;\delta_1(\cdot)]\right)},$$

where $x'\delta\left(F[\eta|x;\delta(\cdot)]\right) = \eta$, and $F[\eta|x;\delta(\cdot)]$ is well-defined for all $\xi(\cdot) \in \mathcal{H}$.

Hence:

$$\left| f_{\eta|X;\delta_2(\cdot)}(\eta|x) - f_{\eta|X;\delta_1(\cdot)}(\eta|x) \right| \leq \frac{\|x\| \left\| \nabla\delta_1\left(F[\eta|x;\delta_1(\cdot)]\right) - \nabla\delta_2\left(F[\eta|x;\delta_2(\cdot)]\right) \right\|}{\underline{c}^2}.$$

Now:

$$\begin{aligned} \left\| \nabla\delta_1\left(F[\eta|x;\delta_1(\cdot)]\right) - \nabla\delta_2\left(F[\eta|x;\delta_2(\cdot)]\right) \right\| &\leq \left\| \nabla\delta_1\left(F[\eta|x;\delta_1(\cdot)]\right) - \nabla\delta_2\left(F[\eta|x;\delta_1(\cdot)]\right) \right\| \\ &\quad + \left\| \nabla\delta_2\left(F[\eta|x;\delta_1(\cdot)]\right) - \nabla\delta_2\left(F[\eta|x;\delta_2(\cdot)]\right) \right\|. \end{aligned} \tag{B3}$$

We will bound the two terms in (B3) in turn. To bound the first term we first establish the following result.

Sublemma 1 *Let $\xi_1(\cdot)$ and $\xi_2(\cdot)$ be in $\mathcal{H}$. For any constant $0 < \nu < \frac{1}{2}$ there exists a constant $\widetilde{C} > 0$ such that:*

$$\left\| \nabla\delta_2(\cdot) - \nabla\delta_1(\cdot) \right\|_\infty \leq \widetilde{C} \left\| \delta_2(\cdot) - \delta_1(\cdot) \right\|_\infty^\nu.$$

Proof. Let $g(\tau)$ denote any of the components of $\nabla\delta_2(\tau) - \nabla\delta_1(\tau)$, and let $a = \|\delta_2(\cdot) - \delta_1(\cdot)\|_\infty$. Take:

$$\widetilde{C} > 2^{\left(\frac{5}{2}-\nu\right)} \overline{c}^{(1-\nu)},$$

and suppose that, for some $\tau_0 \in (0,1)$, $|g(\tau_0)| > \widetilde{C} a^\nu$. We are going to verify that this leads to a contradiction.

Without loss of generality we can assume that $g(\tau_0) > \widetilde{C} a^\nu$. We have, for all $\tau \in (0,1)$:

$$\begin{aligned} g(\tau) &> \widetilde{C} a^\nu - |g(\tau) - g(\tau_0)| \\ &\geq \widetilde{C} a^\nu - 2\overline{c}|\tau - \tau_0| \\ &\geq \frac{\widetilde{C}}{2} a^\nu, \end{aligned}$$

provided $|\tau - \tau_0| \leq \frac{\widetilde{C}}{4\overline{c}} a^\nu$, where we have used that $\xi_1(\cdot)$ and $\xi_2(\cdot)$ belong to $\mathcal{H}$.

Hence, for all τ such that $|\tau - \tau_0| \leq \frac{\widetilde{C}}{4\overline{c}} a^\nu$:

$$\left| \int_{\tau_0}^{\tau} g(u)du \right| \geq \frac{\widetilde{C}}{2} a^\nu |\tau - \tau_0|,$$

from which it follows that:

$$\left| \int_{\tau_0}^{\tau} g(u)du \right| \geq \frac{\widetilde{C}}{2} a^\nu \times \frac{\widetilde{C}}{4\overline{c}} a^\nu = \frac{\widetilde{C}^2}{8\overline{c}} a^{2\nu}.$$

Now, by the definition of $a = \|\delta_2(\cdot) - \delta_1(\cdot)\|_\infty$ we also have:

$$\left| \int_{\tau_0}^{\tau} g(u)du \right| \leq 2a.$$

Noting that $a \leq 2\overline{c}$, this leads to a contradiction and ends the proof of Sublemma 1.

■

Now, Sublemma 1 implies that:

$$\|\nabla\delta_1 (F[\eta|x;\delta_1(\cdot)]) - \nabla\delta_2 (F[\eta|x;\delta_1(\cdot)])\| \leq \widetilde{C} \|\delta_2(\cdot) - \delta_1(\cdot)\|_\infty^\nu .$$

As for the second term in (B3), we have:

$$\|\nabla\delta_2 (F[\eta|x;\delta_1(\cdot)]) - \nabla\delta_2 (F[\eta|x;\delta_2(\cdot)])\| \leq \overline{c} |F[\eta|x;\delta_1(\cdot)] - F[\eta|x;\delta_2(\cdot)]| .$$

Moreover:

$$x'\delta_2 (F[\eta|x;\delta_2(\cdot)]) - x'\delta_1 (F[\eta|x;\delta_1(\cdot)]) = 0,$$

so:

$$\begin{aligned} \left|x'\delta_2 (F[\eta|x;\delta_2(\cdot)]) - x'\delta_2 (F[\eta|x;\delta_1(\cdot)])\right| &= \left|x'\delta_1 (F[\eta|x;\delta_1(\cdot)]) - x'\delta_2 (F[\eta|x;\delta_1(\cdot)])\right| \\ &\leq \|x\| \|\delta_2(\cdot) - \delta_1(\cdot)\|_\infty . \end{aligned}$$

Hence, using the fact that $\xi_2(\cdot) \in \mathcal{H}$:

$$|F[\eta|x;\delta_1(\cdot)] - F[\eta|x;\delta_2(\cdot)]| \leq \frac{1}{\underline{c}} \|x\| \|\delta_2(\cdot) - \delta_1(\cdot)\|_\infty .$$

This shows that the second term in (B3) is suitably bounded. Repeating the argument for the T period-specific likelihood functions, it can then be shown that $I_2 \leq \widetilde{C}_2 \|\xi_2(\cdot) - \xi_1(\cdot)\|_\infty^\nu$.

Finally, using similar arguments it can be shown that:

$$\begin{aligned} II \equiv & \left\| \int X_i \psi_{\tau_1} \left(\eta - X_i'\delta_2(\tau_1)\right) f_i(\eta;\xi_2(\cdot)) d\eta \right. \\ & \left. - \int X_i \psi_{\tau_1} \left(\eta - X_i'\delta_1(\tau_1)\right) f_i(\eta;\xi_1(\cdot)) d\eta \right\| \\ \leq & \widetilde{C}_2 \|\xi_2(\cdot) - \xi_1(\cdot)\|_\infty^\nu , \end{aligned}$$

possibly with a different constant $\widetilde{C}_2$.

This completes the proof of Lemma B1.

■

Lemma B2 *For all* $\xi_1(\cdot) \in \mathcal{H}$ *there is* $\xi_2(\cdot) \in \mathcal{H}^{L_N}$ *such that*

$$|Q_N(\xi_2(\cdot)) - Q_N(\xi_1(\cdot))| = o_p(1). \tag{B4}$$

Moreover, for all $\xi_1(\cdot) \in \mathcal{H}^{L_N}$ *there is* $\xi_2(\cdot) \in \mathcal{H}$ *such that*

$$\sup_{\tau \in (\tau_1, \tau_L)} \|\xi_2(\tau) - \xi_1(\tau)\| = o(1), \quad \text{and} \quad |Q_N(\xi_2(\cdot)) - Q_N(\xi_1(\cdot))| = o_p(1). \tag{B5}$$

Proof.

To each function $\xi(\cdot) : (0,1) \to \mathbb{R}^p$ we associate the interpolating spline $\pi_{L_N}\xi(\cdot)$ that satisfies $\pi_{L_N}\xi(\tau_\ell) = \xi(\tau_\ell)$ for all $\ell \in \{1, ..., L_N\}$, that is piecewise-linear on (τ_1, τ_{L_N}), and satisfies Assumption 5 (vi) on $(0, \tau_1)$ and $(\tau_{L_N}, 1)$.

We start with a preliminary sublemma.

Sublemma 2 *For all* $\xi(\cdot) \in \mathcal{H}$*:*

$$\sup_{\tau \in (\tau_1, \tau_{L_N})} \|\pi_{L_N}\xi(\tau) - \xi(\tau)\| = o(1), \quad \text{and} \quad |Q_N(\pi_{L_N}\xi(\cdot)) - Q_N(\xi(\cdot))| = o_p(1). \tag{B6}$$

Proof.

To prove the first part in Sublemma 2, note that, using that $\pi_{L_N}\xi(\tau_\ell) = \xi(\tau_\ell)$ and the triangle inequality:

$$\begin{aligned}
\sup_{\tau \in (\tau_1, \tau_{L_N})} \|\pi_{L_N}\xi(\tau) - \xi(\tau)\| \leq & \sup_{\ell \in \{1,...,L_N-1\}} \sup_{\tau \in (\tau_\ell, \tau_{\ell+1})} \|\pi_{L_N}\xi(\tau) - \pi_{L_N}\xi(\tau_\ell)\| \\
& + \sup_{\ell \in \{1,...,L_N-1\}} \sup_{\tau \in (\tau_\ell, \tau_{\ell+1})} \|\xi(\tau) - \xi(\tau_\ell)\| \\
\leq & \sup_{\ell \in \{1,...,L_N-1\}} \sup_{\tau \in (\tau_\ell, \tau_{\ell+1})} |\tau - \tau_\ell| \frac{\|\xi(\tau_{\ell+1}) - \xi(\tau_\ell)\|}{|\tau_{\ell+1} - \tau_\ell|} \\
& + \sup_{\ell \in \{1,...,L_N-1\}} \sup_{\tau \in (\tau_\ell, \tau_{\ell+1})} \overline{c}|\tau - \tau_\ell|.
\end{aligned}$$

Hence, by Assumption 6 (iii) and the fact that $\xi(\cdot) \in \mathcal{H}$:

$$\sup_{\tau \in (\tau_1, \tau_{L_N})} \|\pi_{L_N}\xi(\tau) - \xi(\tau)\| \leq 2\overline{c} \sup_{\ell \in \{1,...,L_N-1\}} |\tau_{\ell+1} - \tau_\ell| = o(1).$$

We now prove the second part in Sublemma 2. We have:

$$Q_N(\pi_{L_N}\xi(\cdot)) - Q_N(\xi(\cdot)) = \frac{1}{L_N} \sum_{\ell=1}^{L_N} \left\|\widehat{\mathbb{E}}[\Psi_i(\pi_{L_N}\xi(\cdot), \tau_\ell)]\right\|^2 - \left\|\widehat{\mathbb{E}}[\Psi_i(\xi(\cdot), \tau_\ell)]\right\|^2.$$

As in the proof of Lemma B1, it suffices to bound

$$\begin{aligned}
\|\Psi_i(\pi_{L_N}\xi(\cdot), \tau_\ell) - \Psi_i(\xi(\cdot), \tau_\ell)\| = & \Big\| \int \sum_{t=1}^T W_{it}(\eta) \psi_{\tau_\ell}\left(Y_{it} - W_{it}(\eta)'\theta(\tau_\ell)\right) f_i(\eta; \pi_{L_N}\xi(\cdot))\, d\eta \\
& - \int \sum_{t=1}^T W_{it}(\eta) \psi_{\tau_\ell}\left(Y_{it} - W_{it}(\eta)'\theta(\tau_\ell)\right) f_i(\eta; \xi(\cdot))\, d\eta \Big\| \\
& + \Big\| \int X_i \psi_{\tau_\ell}\left(\eta - X_i'\delta(\tau_\ell)\right) f_i(\eta; \pi_{L_N}\xi(\cdot))\, d\eta \\
& - \int X_i \psi_{\tau_\ell}\left(\eta - X_i'\delta(\tau_\ell)\right) f_i(\eta; \xi(\cdot))\, d\eta \Big\|,
\end{aligned}$$

where we have used that $\pi_{L_N}\xi(\tau_\ell)=\xi(\tau_\ell)$.

We now bound the first term (the second term being similar):

$$\begin{aligned} I &\equiv \left\| \int \sum_{t=1}^{T} W_{it}(\eta)\,\psi_{\tau_\ell}\left(Y_{it}-W_{it}(\eta)'\theta(\tau_\ell)\right)\left[f_i(\eta;\pi_{L_N}\xi(\cdot))-f_i(\eta;\xi(\cdot))\right]d\eta \right\| \\ &\leq C\int \left|f_i(\eta;\pi_{L_N}\xi(\cdot))-f_i(\eta;\xi(\cdot))\right|d\eta, \end{aligned}$$

where $C>0$ is a constant.

Let us define:

$$\begin{aligned} \mathcal{E}_i &= \Bigg\{\eta,\, \eta\leq X_i'\delta(\tau_1)\,,\text{ or } \eta>X_i'\delta(\tau_{L_N})\,,\text{ or } \exists t \text{ s.t. } Y_{it}\leq W_{it}(\eta)'\delta(\tau_1)\,, \\ &\qquad\qquad \text{or } \exists t \text{ s.t. } Y_{it}>W_{it}(\eta)'\delta(\tau_{L_N})\Bigg\}, \end{aligned}$$

and let $\mathcal{E}_i^c$ be its complement.

We have:

$$\begin{aligned} I &\leq C\int_{\mathcal{E}_i^c}\left|f_i(\eta;\pi_{L_N}\xi(\cdot))-f_i(\eta;\xi(\cdot))\right|d\eta + C\int_{\mathcal{E}_i} f_i(\eta;\pi_{L_N}\xi(\cdot))\,d\eta + C\int_{\mathcal{E}_i} f_i(\eta;\xi(\cdot))\,d\eta \\ &\equiv I_1+I_2+I_3. \end{aligned}$$

To bound I_1, as in the proof of Lemma B1 we will bound the difference between the random-effects distributions:

$$\sup_{\eta\in\mathcal{E}_i^c}\left|f_{\eta|X;\pi_{L_N}\delta(\cdot)}(\eta|x)-f_{\eta|X;\delta(\cdot)}(\eta|x)\right|,$$

as similar arguments allow to bound differences between the T period-specific likelihood functions.

Let $\eta=x'\delta(\tau)$ such that $\tau\in(\tau_\ell,\tau_{\ell+1})$ for some $\ell\in\{1,...,L_N-1\}$. We have, using the mean value theorem:

$$\begin{aligned} \left|f_{\eta|X;\pi_{L_N}\delta(\cdot)}(\eta|x)-f_{\eta|X;\delta(\cdot)}(\eta|x)\right| &= \left|\frac{\tau_{\ell+1}-\tau_\ell}{x'\left[\delta(\tau_{\ell+1})-\delta(\tau_\ell)\right]}-f_{\eta|X;\delta(\cdot)}(\eta|x)\right| \\ &= \left|f_{\eta|X;\delta(\cdot)}\left(x'\delta(\widetilde{\tau})|x\right)-f_{\eta|X;\delta(\cdot)}\left(x'\delta(\tau)|x\right)\right|, \end{aligned}$$

where $\widetilde{\tau}\in(\tau_\ell,\tau_{\ell+1})$.

Now, as $\xi(\cdot)\in\mathcal{H}$ it is easy to verify that, for some constant $\widetilde{C}$:

$$\left|f_{\eta|X;\delta(\cdot)}\left(x'\delta(\widetilde{\tau})|x\right)-f_{\eta|X;\delta(\cdot)}\left(x'\delta(\tau)|x\right)\right|\leq\widetilde{C}\left|\widetilde{\tau}-\tau\right|,$$

which is $o_p(1)$, uniformly in ℓ, $\xi(\cdot)$ and η. Repeating the argument for the other components of the posterior distributions shows that $I_1=o_p(1)$.

To complete the proof of Sublemma 2, we need to bound I_2 and I_3. We note that the random-effects density and likelihood functions are bounded from below and above. Hence:

$$\begin{aligned} I_2 &= C\int_{\mathcal{E}_i} f_i(\eta;\pi_{L_N}\xi(\cdot))\,d\eta \\ &\leq \widetilde{C}\int_{-\infty}^{X_i'\delta(\tau_1)} f_{\eta|X;\pi_{L_N}\delta(\cdot)}(\eta|X_i)\,d\eta+\widetilde{C}\int_{X_i'\delta(\tau_{L_N})}^{+\infty} f_{\eta|X;\pi_{L_N}\delta(\cdot)}(\eta|X_i)\,d\eta \\ &= \widetilde{C}\tau_1+\widetilde{C}\left(1-\tau_{L_N}\right), \end{aligned}$$

for some constant $\widetilde{C} > 0$, where we have used Assumption 5 (vi). Hence $I_2 = o_p(1)$, uniformly.

Similarly, for some (possibly different) constant $\widetilde{C} > 0$:

$$\begin{aligned} I_3 &= C \int_{\mathcal{E}_i} f_i\left(\eta; \xi(\cdot)\right) d\eta \\ &\leq \widetilde{C} \int_{-\infty}^{X_i'\delta(\tau_1)} f_{\eta|X;\delta(\cdot)}\left(\eta|X_i\right) d\eta + \widetilde{C} \int_{X_i'\delta(\tau_{L_N})}^{+\infty} f_{\eta|X;\delta(\cdot)}\left(\eta|X_i\right) d\eta \\ &= \widetilde{C}\tau_1 + \widetilde{C}\left(1 - \tau_{L_N}\right), \end{aligned}$$

by definition of $\delta(\cdot)$. Hence $I_3 = o_p(1)$, uniformly.

This ends the proof of Sublemma 2.

■

To complete the proof of Lemma B2, let $\xi_1(\cdot) \in \mathcal{H}$. We normalize $\beta_0(\cdot)$ and $\gamma(\cdot)$ in $\xi_1(\cdot)$ so that Assumption 5 (iii) holds. By Assumption 4 (ii) this condition holds asymptotically as N and L_N tend to infinity. Then it can be shown that $\pi_{L_N}\xi(\cdot) \in \mathcal{H}^{L_N}$, possibly by taking larger (resp., lower) $\overline{c}$ and $\underline{c}$ in Assumption 5. Then by Sublemma 2 the first part of Lemma B2 holds true.

To show the second part of Lemma B2, let $\xi_1(\cdot) \in \mathcal{H}^{L_N}$. Consider a smooth interpolating spline $\xi_2(\cdot) \in \mathcal{H}$ that satisfies $\xi_2(\tau_\ell) = \xi_1(\tau_\ell)$ for all $\ell = 1, ..., L_N$, possibly taking larger (resp., lower) $\overline{c}$ and $\underline{c}$ in Assumption 4. We impose the mean normalization of Assumption 4 (ii). Then, by construction, $\xi_1(\cdot) = \pi_{L_N}\xi_2(\cdot)$, so the second part of Lemma B2 follows from Sublemma 2.

This ends the proof of Lemma B2.

■

We now prove Proposition 2. By Assumption 6 (ii), $\widehat{\xi}(\cdot) \in \mathcal{H}^{L_N}$ w.p.a.1. Hence by Lemma B2 there is $\widetilde{\xi}(\cdot) \in \mathcal{H}$ such that:

$$\sup_{\tau \in (\tau_1, \tau_L)} \left\| \widehat{\xi}(\tau) - \widetilde{\xi}(\tau) \right\| = o_p(1), \tag{B7}$$

and:

$$\left| Q_N\left(\widehat{\xi}(\cdot)\right) - Q_N\left(\widetilde{\xi}(\cdot)\right) \right| = o_p(1). \tag{B8}$$

Similarly, $\xi^0(\cdot) \in \mathcal{H}$, so by Lemma B2 there is $\widetilde{\xi}^0(\cdot) \in \mathcal{H}^{L_N}$ such that:[24]

$$\left| Q_N\left(\xi^0(\cdot)\right) - Q_N\left(\widetilde{\xi}^0(\cdot)\right) \right| = o_p(1). \tag{B9}$$

Moreover, by Assumption 6 (ii):

$$Q_N\left(\widehat{\xi}(\cdot)\right) \leq Q_N\left(\widetilde{\xi}^0(\cdot)\right) + o_p(1),$$

so, by (B8) and (B9):

$$Q_N\left(\widetilde{\xi}(\cdot)\right) \leq Q_N\left(\xi^0(\cdot)\right) + o_p(1). \tag{B10}$$

Next, we apply Lemma A.2 in Newey and Powell (2003), which we state here for completeness.

Lemma B3 *(Newey and Powell, 2003) Let $\mathcal{H}$ be compact with respect to $\|\cdot\|_\infty$. Suppose that $Q_N(\xi(\cdot)) - Q(\xi(\cdot)) = o_p(1)$ for all $\xi(\cdot) \in \mathcal{H}$, and that there is $\nu > 0$ and $M_N = O_p(1)$ such that $|Q_N(\xi_2(\cdot)) - Q_N(\xi_1(\cdot))| \leq M_N \|\xi_2(\cdot) - \xi_1(\cdot)\|_\infty^\nu$ for all $\xi_1(\cdot), \xi_2(\cdot) \in \mathcal{H}$. Then $Q(\cdot)$ is continuous on $\mathcal{H}$, and $\sup_{\xi(\cdot)\in\mathcal{H}} |Q_N(\xi(\cdot)) - Q(\xi(\cdot))| = o_p(1)$.*

[24] Note that, in the proof of Lemma B2, $\mathcal{H}$ and $\mathcal{H}^{L_N}$ may correspond to different constants $\overline{c}$ and $\underline{c}$. In this case Assumption 6 needs to be suitably modified. Specifically, letting $\overline{c}_1 < \overline{c}_2 < \overline{c}_3$ and $\underline{c}_1 > \underline{c}_2 > \underline{c}_3$, one may let $\xi^0(\cdot) \in \mathcal{H}_{\overline{c}_1, \underline{c}_1}$, Assumption 6 (i) hold with $\mathcal{H}_{\overline{c}_3, \underline{c}_3}$, and Assumption 6 (ii) hold with $\mathcal{H}^{L_N}_{\overline{c}_2, \underline{c}_2}$.

Let us suppose to start with that the conditions of Lemma B3 are satisfied. Then by (B10) and Lemma B3 we obtain:

$$Q\left(\widetilde{\xi}(\cdot)\right) \leq Q\left(\xi^0(\cdot)\right) + o_p(1). \tag{B11}$$

Now, by Lemma B3 $Q(\cdot)$ is continuous on $\mathcal{H}$, which is is compact in the topology of the supremum norm $\|\cdot\|_\infty$, as it is a closed subset of a compact Hölder ball; see for example Gallant and Nychka (1987) for detail. It thus follows from Assumption 6 (i) that, for all $\epsilon > 0$:

$$\left[\inf_{\xi(\cdot)\in\mathcal{H},\left\|\xi(\cdot)-\xi^0(\cdot)\right\|_\infty \geq \epsilon} \ Q(\xi(\cdot))\right] \ > \ Q(\xi^0(\cdot)).$$

Hence, by (B11):

$$\left\|\widetilde{\xi}(\cdot) - \xi^0(\cdot)\right\|_\infty = o_p(1).$$

Combining with (B7) and using the triangle inequality, it follows:

$$\sup_{\tau\in(\tau_1,\tau_L)} \left\|\widehat{\xi}(\tau) - \xi^0(\tau)\right\| = o_p(1).$$

Hence Proposition 2 is proven, provided it can be shown that the conditions of Lemma B3 are satisfied. We first need to verify that, for all $\xi(\cdot) \in \mathcal{H}$:

$$Q_N\left(\xi(\cdot)\right) \overset{p}{\to} Q\left(\xi(\cdot)\right). \tag{B12}$$

To see this, note that we have, for all $\xi(\cdot) \in \mathcal{H}$:

$$\begin{aligned} Q_N\left(\xi(\cdot)\right) - Q\left(\xi(\cdot)\right) \quad &= \quad \frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\widehat{\mathbb{E}}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 - \frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 \\ &\quad + \frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 - \int_0^1\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau\right)\right]\right\|^2 d\tau \\ &\equiv \quad A_N + B_N. \end{aligned}$$

We are going to show that $A_N = o_p(1)$ and $B_N = o(1)$. We start by noting that, for all $\epsilon > 0$:

$$\begin{aligned} \Pr\left(|A_N| > \epsilon\right) \quad &= \quad \Pr\left(\left|\frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\widehat{\mathbb{E}}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 - \frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2\right| > \epsilon\right) \\ &\leq \quad \Pr\left(\sup_{\ell\in\{1,\ldots,L_N\}}\left|\left\|\widehat{\mathbb{E}}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 - \left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2\right| > \epsilon\right) \\ &\leq \quad L_N\sup_{\ell\in\{1,\ldots,L_N\}}\Pr\left(\left|\left\|\widehat{\mathbb{E}}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 - \left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2\right| > \epsilon\right), \end{aligned}$$

where we have used the union bound.

Then it is easy to see that, by Assumption 6 (iv):

$$\begin{aligned} \sup_{\ell\in\{1,\ldots,L_N\}}\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\| \quad &= \quad O(1), \\ \sup_{\ell\in\{1,\ldots,L_N\}}\left\|\widehat{\mathbb{E}}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\| \quad &= \quad O_p(1). \end{aligned}$$

Moreover, by the Chebyshev inequality we have, for all $\ell \in \{1, ..., L_N\}$:

$$L_N \Pr\left(\left|\left\|\widehat{\mathbb{E}}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\| - \left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|\right| > \epsilon\right) \quad \leq \quad \frac{L_N}{N}\operatorname{Var}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right],$$

where $\sup_{\ell\in\{1,\ldots,L_N\}} \mathrm{Var}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]$ is bounded by Assumption 6 (*iv*). It thus follows that $A_N = o_p(1)$.

Turning to B_N we have:

$$\begin{aligned} B_N &= \frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 - \int_0^1 \left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau\right)\right]\right\|^2 d\tau \\ &= \frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 - \sum_{\ell=1}^{L_N}\left(\tau_{\ell+1}-\tau_\ell\right)\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 \\ &\quad + \sum_{\ell=1}^{L_N}\left(\tau_{\ell+1}-\tau_\ell\right)\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 - \int_0^1 \left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau\right)\right]\right\|^2 d\tau \\ &= \sum_{\ell=1}^{L_N}\left[\left(\tau_{\ell+1}-\tau_\ell\right)\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 - \int_{\tau_\ell}^{\tau_{\ell+1}} \left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau\right)\right]\right\|^2 d\tau\right] \\ &\quad - \int_0^{\tau_1} \left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau\right)\right]\right\|^2 d\tau + o(1), \end{aligned}$$

where we have used Assumption 6 (*iii*) and (*iv*). Note that, by the same assumptions:

$$\int_0^{\tau_1} \left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau\right)\right]\right\|^2 d\tau = o(1).$$

It thus follows that:

$$\begin{aligned} B_N &= \sum_{\ell=1}^{L_N}\int_{\tau_\ell}^{\tau_{\ell+1}}\left[\left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau_\ell\right)\right]\right\|^2 - \left\|\mathbb{E}\left[\Psi_i\left(\xi(\cdot),\tau\right)\right]\right\|^2\right] d\tau + o(1) \\ &\leq C_1\sum_{\ell=1}^{L_N}\int_{\tau_\ell}^{\tau_{\ell+1}} \left|\tau-\tau_\ell\right| d\tau. \end{aligned}$$

where we have used Assumption 6 (*iv*) and Lemma B1.

Hence, by Assumption 6 (*iii*):

$$B_N \leq C_1\sum_{\ell=1}^{L_N}\left|\tau_{\ell+1}-\tau_\ell\right|^2 = o(1).$$

To complete the proof, we need to verify the last condition in Lemma B3. This follows from Lemma B1, as, for all $(\xi_1(\cdot),\xi_2(\cdot)) \in \mathcal{H}\times\mathcal{H}$:

$$\begin{aligned} \left|Q_N\left(\xi_2(\cdot)\right) - Q_N\left(\xi_1(\cdot)\right)\right| &= \left|\frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\widehat{\mathbb{E}}\left[\Psi_i\left(\xi_2(\cdot),\tau_\ell\right)\right]\right\|^2 - \frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\widehat{\mathbb{E}}\left[\Psi_i\left(\xi_1(\cdot),\tau_\ell\right)\right]\right\|^2\right| \\ &\leq \left(\frac{1}{L_N}\sum_{\ell=1}^{L_N}\left\|\widehat{\mathbb{E}}\left[\Psi_i\left(\xi_2(\cdot),\tau_\ell\right) - \Psi_i\left(\xi_1(\cdot),\tau_\ell\right)\right]\right\|\right) \times O_p(1) \\ &\leq \left(C_2\left\|\xi_2(\cdot)-\xi_1(\cdot)\right\|_\infty^\nu\right) \times O_p(1) \\ &\leq O_p(1)\left\|\xi_2(\cdot)-\xi_1(\cdot)\right\|_\infty^\nu. \end{aligned}$$

This completes the verification of the conditions in Lemma B3, and thus ends the proof of Proposition 2.

C Nonparametric identification: sketch of the arguments

We consider two setups in turn: the static model of Section 2, and the first-order Markov model of Section 6.1. In both cases we provide a brief outline of the identification argument.

Static model. Let us start with the static model of Section 2. Building on Hu and Schennach (2008), we define the following linear operators, which act on spaces of bounded functions.

Let y_2 be one element in the support of Y_2. To a function $h : y_1 \mapsto h(y_1)$ we associate:

$$L_{Y_1,\eta}h : \eta \mapsto \int f_{Y_1,\eta}(y_1,\eta)h(y_1)dy_1,$$

and

$$L_{Y_1,(y_2),Y_3}h : y_3 \mapsto \int f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3)h(y_1)dy_1.$$

To a function $g : \eta \mapsto g(\eta)$ we associate:

$$\Delta_{(y_2)|\eta}g : \eta \mapsto f_{Y_2|\eta}(y_2|\eta)g(\eta),$$

and

$$L_{Y_3|\eta}g : y_3 \mapsto \int f_{Y_3|\eta}(y_3|\eta)g(\eta)d\eta.$$

We have, for all functions $h : y_1 \mapsto h(y_1)$:

$$\begin{aligned}
[L_{Y_1,(y_2),Y_3}h](y_3) &= \int f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3)h(y_1)dy_1 \\
&= \int \left[\int f_{Y_3|\eta}(y_3|\eta)f_{Y_2|\eta}(y_2|\eta)f_{Y_1,\eta}(y_1,\eta)d\eta\right] h(y_1)dy_1 \\
&= \int f_{Y_3|\eta}(y_3|\eta)f_{Y_2|\eta}(y_2|\eta)\left[\int f_{Y_1,\eta}(y_1,\eta)h(y_1)dy_1\right] d\eta \\
&= [L_{Y_3|\eta}\Delta_{(y_2)|\eta}L_{Y_1,\eta}h](y_3).
\end{aligned}$$

We thus have:

$$L_{Y_1,(y_2),Y_3} = L_{Y_3|\eta}\Delta_{(y_2)|\eta}L_{Y_1,\eta}, \quad \text{for all } y_2. \tag{C13}$$

This yields a joint diagonalization system of operators, because, under suitable invertibility (i.e., injectivity) conditions, (C13) implies:

$$L_{Y_1,(y_2),Y_3}L^{-1}_{Y_1,(\widetilde{y}_2),Y_3} = L_{Y_3|\eta}\Delta_{(y_2)|\eta}\Delta^{-1}_{(\widetilde{y}_2)|\eta}L^{-1}_{Y_3|\eta}, \quad \text{for all } y_2, \widetilde{y}_2. \tag{C14}$$

The conditions of Hu and Schennach (2008)'s theorem then guarantee uniqueness of the solutions to (C14).

Dynamic autoregressive model. Let us now consider the dynamic autoregressive setup of Section 6.1. As in Hu and Shum (2012) we can define the following operators.

Let (y_2, y_3) be an element in the support of (Y_2, Y_3). To a function $h : y_1 \mapsto h(y_1)$ we associate:

$$L_{Y_1,(y_2),\eta}h : \eta \mapsto \int f_{Y_1,Y_2,\eta}(y_1,y_2,\eta)h(y_1)dy_1,$$

and

$$L_{Y_1,(y_2),(y_3),Y_4} h : y_4 \mapsto \int f_{Y_1,Y_2,Y_3,Y_4}(y_1, y_2, y_3, y_4) h(y_1) dy_1.$$

To a function $g : \eta \mapsto g(\eta)$ we associate:

$$\Delta_{(y_3)|(y_2),\eta} g : \eta \mapsto f_{Y_3|Y_2,\eta}(y_3|y_2, \eta) g(\eta),$$

and

$$L_{Y_4|(y_3),\eta} g : y_4 \mapsto \int f_{Y_4|(y_3),\eta}(y_4|y_3, \eta) g(\eta) d\eta.$$

As above we verify that:

$$L_{Y_1,(y_2),(y_3),Y_4} = L_{Y_4|(y_3),\eta} \Delta_{(y_3)|(y_2),\eta} L_{Y_1,(y_2),\eta}, \quad \text{for all } (y_2, y_3). \tag{C15}$$

Hence, under suitable invertibility conditions:

$$L_{Y_1,(y_2),(y_3),Y_4} L^{-1}_{Y_1,(\widetilde{y}_2),(\widetilde{y}_3),Y_4} = L_{Y_4|(y_3),\eta} \Delta_{(y_3)|(y_2),\eta} \Delta^{-1}_{(\widetilde{y}_3)|(\widetilde{y}_2),\eta} L^{-1}_{Y_4|(\widetilde{y}_3),\eta}, \quad \text{for all } y_2, \widetilde{y}_2, y_3, \widetilde{y}_3. \tag{C16}$$

Hu and Shum (2012)—in particular in their Lemma 3—provide conditions for uniqueness of the solutions to (C16). Their conditions are closely related to the ones in Hu and Schennach (2008); see Assumption 3 above.