Nonlinear Panel Data Estimation via Quantile Regressions*

Manuel Arellano
CEMFI, Madrid

Stéphane Bonhomme
University of Chicago

First draft: February 2013
Revised draft: January 2016

Abstract

We introduce a class of quantile regression estimators for short panels. Our framework covers static and dynamic autoregressive models, models with general predetermined regressors, and models with multiple individual effects. We use quantile regression as a flexible tool to model the relationships between outcomes, covariates, and heterogeneity. We develop an iterative simulation-based approach for estimation, which exploits the computational simplicity of ordinary quantile regression in each iteration step. Finally, an application to measure the effect of smoking during pregnancy on birthweight completes the paper.

JEL code: C23.
Keywords: Panel data, dynamic models, non-separable heterogeneity, quantile regression, expectation-maximization.

*A presentation based on this paper was given at the *Econometrics Journal* RES Conference 2013. A previous version of the paper was called: "Random-Effects Quantile Regression". We thank the editor and two anonymous referees for comments and suggestions. We also thank Richard Blundell, Xiaohong Chen, Andrew Chesher, Bryan Graham, Roger Koenker, Arthur Lewbel, 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, the literature is still short of answers for panel versions of many models commonly used in empirical work (Arellano and Bonhomme, 2011). More broadly, to date no approach is yet available to specify and estimate general panel data relationships in static or dynamic settings.

In this paper we rely on quantile regression as a flexible estimation tool for nonlinear panel models. Since Koenker and Bassett (1978), quantile regression techniques have proven useful tools to document distributional effects in cross-sectional settings. Koenker (2005) provides a comprehensive account of these methods. Quantile-based specifications have the ability to deal with complex interactions between covariates and latent heterogeneity, and to provide a rich description of heterogeneous responses of outcomes to variations in covariates. In panel data, quantile methods are particularly well-suited as they allow building flexible models for the dependence of unobserved heterogeneity on exogenous covariates or initial conditions, and for the feedback processes of covariates in dynamic models with general predetermined regressors.

We consider classes of panel data models with continuous outcomes that satisfy conditional independence restrictions. In static settings, these conditions restrict the time-series dependence of the time-varying disturbances. 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). In dynamic settings, finite-order Markovian setups naturally imply conditional independence restrictions. In both static and dynamic settings, results from the literature on nonlinear measurement error models (Hu and Schennach, 2008, Hu and Shum, 2012) can then be used to provide sufficient conditions for nonparametric identification for a fixed number of time periods.

The main goal of the paper is to develop a tractable estimation strategy for nonlinear panel models. For this purpose, we specify outcomes Y_{it} as a function of covariates X_{it} and latent heterogeneity η_i as:

$$Y_{it} = \sum_{k=1}^{K_1} \theta_k(U_{it}) g_k \left(X_{it}, \eta_i \right), \tag{1}$$

and we similarly specify the dependence of η_i on covariates $X_i = (X_{i1}', ..., X_{iT}')'$ as:

$$\eta_i = \sum_{k=1}^{K_2} \delta_k(V_i) h_k \left(X_i\right), \tag{2}$$

where $U_{i1}, ..., U_{iT}, V_i$ are independent uniform random variables, and g's and h's belong to some families of functions. Outcomes Y_{it} and heterogeneity η_i are monotone in U_{it} and V_i, respectively, so (1) and (2) are models of conditional quantile functions.

The g's and h's are anonymous functions without an economic interpretation. They are just building blocks of flexible models. Objects of interest will be summary measures of derivative effects constructed from the models.

The linear quantile specifications (1) and (2) allow documenting interactions between covariates and heterogeneity at various quantiles. In particular, (2) is a correlated random-effects model that can become arbitrarily flexible as K_2 increases. Linearity in the quantile parameters, though not essential to our approach, is helpful for computational purposes. Moreover, while (1) and (2) are stated for the static case and a scalar unobserved effect, we show how to extend the framework to allow for dynamics and multi-dimensional latent components.

The main econometric challenge is that the researcher has no data on heterogeneity η_i. If η_i were observed, one would simply run an ordinary quantile regression of Y_{it} on the $g_k(X_{it}, \eta_i)$ variables in (1). As η_i is not observed we need to construct some imputations, say M imputed values $\eta_i^{(m)}$, $m = 1, ..., M$, for each individual in the panel. Having got those, we can get estimates by running a quantile regression averaged over imputed values.

For the imputed values to be valid they have to be draws from the posterior distribution of η_i conditioned on the data, which depends on the parameters to be estimated (θ's and δ's). Our approach is thus iterative. We start by selecting initial values for conditional quantiles of Y_{it} and η_i, which then allows us to generate imputes of η_i, which we can use to update the quantile parameter estimates, and so on.

A difficulty for applying this idea is that the unknown parameters θ's and δ's are functions, hence infinite-dimensional. This is because we need to model the full conditional distribution of outcomes and latent individual effects, as opposed to a single quantile as is typically the case in applications of ordinary quantile regression. To deal with this issue we follow Wei and Carroll (2009), and we use a finite-dimensional approximation to θ's and δ's based on interpolating splines.

In the case of model (1) and (2), the estimation method works as follows, starting with initial parameter values for $\theta_k(\tau)$ and $\delta_k(\tau)$ and iterating the two steps below until convergence to a stationary distribution:

1. Given values for $\theta_k(\tau)$ and $\delta_k(\tau)$ on a grid of τ's, we compute the implied posterior distribution of the individual effects and draw, for each individual unit i in the sample, a sequence $\eta_i^{(1)}, ..., \eta_i^{(M)}$ from that distribution.

2. With draws of η's at hand, we update the parameters $\theta_k(\tau)$ and $\delta_k(\tau)$ by means of two sets of quantile regressions, regressing outcomes Y_{it} on the $g_k(X_{it}, \eta_i^{(m)})$ to update $\theta_k(\tau)$, and regressing the individual draws $\eta_i^{(m)}$ on the $h_k(X_i)$ to update $\delta_k(\tau)$.

The resulting algorithm is a variant of the Expectation-Maximization algorithm of Dempster, Laird and Rubin (1977), sometimes referred to as "stochastic EM". The sequence of parameter estimates converges to an ergodic Markov Chain in the limit. Following Nielsen (2000a, 2000b) we characterize the asymptotic distribution of our sequential simulated method-of-moments estimator based on M imputations. A difference with most applications of EM-type algorithms is that we do not update parameters in each iteration using maximum likelihood, but using quantile regressions.[1] This is an important feature of our approach, as the fact that quantile regression estimates can be computed in a quantile-by-quantile fashion, and the convexity of the quantile regression objective function, make each parameter update in step 2 fast and reliable.

We apply our estimator to assess the effect of smoking during pregnancy on a child's birthweight. 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, and exhibits substantial heterogeneity across mothers.

Literature review and outline. Starting with Koenker (2004), most panel data approaches to date proceed in a quantile-by-quantile fashion, and include individual indicators

[1]Related sequential method-of-moments estimators are considered in Arcidiacono and Jones (2003), Arcidiacono and Miller (2011), and Bonhomme and Robin (2009), among others. 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.

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) and Arellano and Weidner (2015) study the large N, T properties of the fixed-effects quantile regression estimator, and show that it may suffer from large biases in short panels. Rosen (2012) shows that a fixed-effects model for a single quantile may not be point-identified. Recent related contributions are Lamarche (2010), Galvao (2011), and Canay (2011). In contrast, our approach relies on specifying a semi-parametric model for individual effects given covariates and initial conditions, as in (2). As a result, in this paper the analysis is conducted for fixed T, as N tends to infinity.

Our approach is closer in spirit to other random-effects approaches in the literature. For example, Abrevaya and Dahl (2008) consider a correlated random-effects model to study the effects of smoking and prenatal care on birthweight. Their approach mimics control function approaches used in linear panel models. Geraci and Bottai (2007) consider a random-effects approach for a single quantile assuming that the outcome variable is distributed as an asymmetric Laplace distribution conditional on covariates and individual effects. Recent related approaches to quantile panel data models include Chernozhukov *et al.* (2013, 2015) and Graham *et al.* (2015). These approaches are non-nested with ours. In particular, they will generally not recover the quantile effects we focus on in this paper. More broadly, compared to existing work, our aim is to build a framework that can deal with general nonlinear and dynamic relationships, thus providing an extension of standard linear panel data methods to nonlinear settings.

The analysis also relates to method-of-moments estimators for models with latent variables. Compared to Schennach (2014), here we rely on conditional moment restrictions and focus on cases where the entire model specification is point-identified. Finally, 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. 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.

The outline of the paper is as follows. In Section 2 we present static models and discuss identification. In Section 3 we present our estimation method and study some of its properties. In Section 4 we extend the approach to dynamic settings. In Section 5 we show how our method can be used to estimate average marginal effects, which are of interest in a number of applications. In Section 6 we present the empirical illustration. Lastly, we conclude in Section 7. Proofs and further discussion are contained in the appendix. Computer codes implementing the method are available as supplementary material.

2 Quantile models for panel data

In this section we start by introducing a class of static panel data models. At the end of the section we provide conditions for nonparametric identification.

2.1 Model and assumptions

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

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

Model (3) can be used to empirically document nonlinear and heterogeneous effects of covariates. In our illustration to smoking and birthweight, the model allows smoking effects to differ across mothers (through the dependence on η_i) and along the distribution of birthweights (through the dependence on U_{it}). In Section 5 we will describe a set of treatment effect parameters that our method allows us to estimate.

We make the following assumption.

Assumption 1. *(outcomes)*

(i) U_{it} follows a standard uniform distribution, independent of (X_i, η_i).

(ii) $\tau \mapsto Q_Y(x, \eta, \tau)$ is strictly increasing on $(0, 1)$, for almost all (x, η) in the support of (X_{it}, η_i).

(iii) For all $t \neq s$, U_{it} is independent of U_{is}.

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. Strict exogeneity of X's can be relaxed to allow for predetermined covariates, see Section 4. 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 all $\tau \in (0,1)$, $Q_Y\left(X_{it}, \eta_i, \tau\right)$ is the τ-conditional quantile of Y_{it} given (X_i, η_i).[2]

Assumption 1 (iii) imposes independence restrictions on the process $U_{i1}, ..., U_{iT}$. Restricting the dynamics of error variables U_{it} is needed when aiming at separating the time-varying unobserved errors U_{it} from the time-invariant unobserved individual effects η_i. In part (iii), U_{it} are assumed to be independent over time. In Section 4 we develop various extensions of the model that allow for dynamic effects. Finally, although we have assumed in (3) that Q_Y does not depend on time, one could easily allow $Q_Y = Q_Y^t$ to depend on t, reflecting for example age or calendar time effects depending on the application.

Unobserved heterogeneity. Next, we specify the conditional quantile response function of η_i given X_i as follows:

$$\eta_i = Q_\eta\left(X_i, V_i\right), \quad i = 1, ..., N. \tag{4}$$

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, independent of X_i.*

(ii) *$\tau \mapsto Q_\eta\left(x, \tau\right)$ is strictly increasing on $(0,1)$, for almost all x in the support of X_i.*

2.2 Examples

We next describe several examples to illustrate the static setup introduced above.

Example 1: Location-scale. As a first special case of model (3), consider the following panel generalization of the location-scale model (He, 1997):

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

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

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

where ε_{it} are i.i.d. across periods, and independent of all regressors and individual effects.[3] 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: Panel quantile regression. Consider next the following linear quantile specification with scalar η_i, which generalizes (5):

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

Given Assumption 1 (i) and (ii), the conditional quantiles of Y_{it} are given by:

$$Q_Y(X_{it}, \eta_i, \tau) = X_{it}'\beta(\tau) + \eta_i\gamma(\tau).$$

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:

$$Y_{it} = X_{it}'\beta(U_{it}) + \eta_i(U_{it}), \tag{7}$$

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 (6), the presence of the process $\eta_i(\tau)$ in (7) introduces an element of nonparametric functional heterogeneity in the conditional distribution of Y_{it}. In contrast, a key aspect of our approach is that we view the η's as missing data, and introduce them as additional (latent) covariates in the quantile regression model.

The term $\eta_i(U_{it})$ in model (7) can be regarded as a function of U_{it} and a vector of unobserved individual effects of unspecified dimension. In this way model (7) allows for multiple individual characteristics that affect differently individuals with different error rank U_{it}. However, while being agnostic about the number of unobserved individual factors affecting outcomes is attractive, sometimes substantive reasons suggest that only a small number of

[3] A generalization of (5) that allows for two-dimensional individual effects—as in Example 3 below—is:

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

underlying factors play a role. Additionally, as our analysis makes clear, whether one uses a quantile model with a different individual effect at each quantile or a model with a small number of unobserved effects has implications for identification.[4]

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

$$\eta_i = X_i'\delta\left(V_i\right). \tag{8}$$

Given Assumption 2, the conditional quantiles of η_i are then given by:

$$Q_\eta(X_i, \tau) = X_i'\delta\left(\tau\right).$$

Model (8) 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):

$$\eta_i = X_i'\mu + \sigma\varepsilon_i, \quad \varepsilon_i|X_i \sim \mathcal{N}\left(0,1\right). \tag{9}$$

For example, in contrast with (9), model (8) is fully nonparametric in the absence of covariates, i.e., when an independent random-effects specification is assumed. Model (8) and its extensions based on series specifications may also be of interest in other nonlinear panel data models, where the outcome equation does not follow a quantile model. We will return to this point in the conclusion.

Example 3: Multi-dimensional heterogeneity. Model (6) may easily be modified to allow for more general interactions between observables and unobservables, thus permitting the effects of covariates to be heterogeneous at different quantiles. A random coefficients generalization that allows for heterogeneous effects is:

$$Q_Y\left(X_{it}, \eta_i, \tau\right) = X_{it}'\beta(\tau) + \gamma_1(\tau)\eta_{i1} + X_{it}'\gamma_2(\tau)\eta_{i2}, \tag{10}$$

where $\eta_i = (\eta_{i1}, \eta_{i2})'$ is bivariate.

In order to extend (8) to the case with bivariate unobserved heterogeneity, it is convenient to assume a triangular structure such as:

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

[4]As mentioned in the introduction, Rosen (2012) shows that a fixed-effects model for a single quantile may not be point-identified.

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 (11) provides a flexible specification for the bivariate conditional distribution of (η_{i1}, η_{i2}) given X_i.[5]

2.3 Nonparametric identification

The class of panel data models introduced above satisfies conditional independence restrictions, as period-specific outcomes $Y_{i1}, ..., Y_{iT}$ are mutually independent conditional on exogenous covariates and individual heterogeneity X_i, η_i. A body of work, initially developed in the context of nonlinear measurement error models, has established nonparametric identification results in related models under conditional independence restrictions; see Hu (2015) for a recent survey. Here we show how the result in Hu and Schennach (2008) can be used to show nonparametric identification. In Section 4 we will build on Hu and Shum (2012) to provide conditions for identification in dynamic models, under Markovian restrictions.

Consider model (3)-(4), with a scalar unobserved effect η_i. At least three periods are needed for identification, and we set $T = 3$. In the case where η_i is multivariate, identification requires using additional time periods, see below. Throughout we use f_Z and $f_{Z|W}$ as generic notation for the distribution function of a random vector Z and for the conditional distribution of Z given W, respectively.

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:

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

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

$$f_{Y_1,Y_2,Y_3|X}\left(y_1, y_2, y_3 \mid x\right) = \int f_{Y_1|\eta,X}\left(y_1 \mid \eta, x\right) f_{Y_2|\eta,X}\left(y_2 \mid \eta, x\right) f_{Y_3|\eta,X}\left(y_3 \mid \eta, x\right) f_{\eta|X}\left(\eta \mid x\right) d\eta. \tag{13}$$

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_{i3} would be the outcome variable, Y_{i2} would be the mismeasured regressor, Y_{i1} would be the instrumental variable,

[5] It is worth pointing out that quantiles appear not to generalize easily to the multivariate case.

and η_i would be the latent, error-free regressor. We closely rely on their analysis and make the following assumption.

Assumption 3. *(identification)*

Almost surely in covariate values x:

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

(ii) For all $\eta_1 \neq \eta_2$: $\Pr\left[f_{Y_3|\eta,X}(Y_{i3}|\eta_1, x) \neq f_{Y_3|\eta,X}(Y_{i3}|\eta_2, x) \,|\, X_i = x\right] > 0.$

(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 $f_{Y_2|\eta,X=x}$ and $f_{Y_1|Y_2,X=x}$, respectively, are injective.

Part (i) in Assumption 3 requires bounded densities. Part (ii) requires that $f_{Y_3|\eta,X}$ be non-identical at different values of η. Part (iii) imposes a centered measure of location on $f_{Y_2|\eta,X=x}$. In Example 2, the following normalization implies Assumption 3 (iii):

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

where $\beta_0(\tau)$ corresponds to the coefficient of the constant in X_{it}. We will use (14) in our empirical implementation.[6] Lastly, part (iv) is an injectivity condition. As pointed out by Hu and Schennach (2008), injectivity is closely related to completeness conditions commonly assumed in the literature on nonparametric instrumental variables. Similarly as completeness, injectivity is a high-level condition.[7] In Appendix A we further discuss the different parts of Assumption 3.

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

Proposition 1. *(Hu and Schennach, 2008)*

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

[6] In fact, Assumption 3 (iii) is also implied by (14) in the following model with first-order interactions, a version of which we estimate in the empirical application: $Y_{it} = X_{it}'\beta\left(U_{it}\right) + \eta_i X_{it}'\gamma\left(U_{it}\right)$.

[7] See for example Canay *et al.* (2012) for results on the testability of completeness assumptions, and D'Haultfoeuille (2011), Andrews (2011), and Hu and Shiu (2012) for primitive conditions in several settings.

This result places no restrictions on the form of $f_{Y_t|\eta,X=x}$, thus allowing for general distributional time effects.

Lastly, the identification result extends to models with multiple, q-dimensional individual effects η_i, by 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_i using (Y_{i1}, Y_{i2}) instead of Y_{i1}, (Y_{i3}, Y_{i4}) instead of Y_{i2}, and Y_{i5} instead of Y_{i3}. Provided injectivity conditions hold, nonparametric identification follows from similar arguments as in the scalar case.

3 Quantile regression estimators

In this section we introduce our estimation strategy and discuss several of its statistical properties.

3.1 Model specification and moment restrictions

We specify the conditional quantile function of Y_{it} in (3) as:

$$Q_Y\left(X_{it}, \eta_i, \tau\right) = W_{it}\left(\eta_i\right)' \theta\left(\tau\right). \tag{15}$$

In (15) the vector $W_{it}\left(\eta_i\right)$ contains a finite number of functions of X_{it} and η_i. One possibility is to adopt a simple linear quantile specification as in Example 2, in which case $W_{it}\left(\eta_i\right) = (X_{it}', \eta_i)'$. A more flexible approach is to use a series specification of the quantile function as in (1), and to set $W_{it}\left(\eta_i\right) = (g_1(X_{it}, \eta_i), ..., g_{K_1}(X_{it}, \eta_i))'$ for a set of K_1 functions $g_1, ..., g_{K_1}$. In practice one may use orthogonal polynomials, wavelets or splines, for example; see Chen (2007) for a comprehensive survey of sieve methods.

Similarly, we specify the conditional quantile function of η_i in (4) as:

$$Q_\eta\left(X_i, \tau\right) = Z_i'\delta\left(\tau\right). \tag{16}$$

In (16) the vector Z_i contains a finite number of functions of covariates X_i, such as $Z_i = (h_1(X_i), ..., h_{K_2}(X_i))$ for a set of K_2 functions $h_1, ..., h_{K_2}$.

The posterior density of the individual effects $f_{\eta|Y,X}$ plays an important role in the analysis. It is given by:

$$f_{\eta|Y,X}\left(\eta \mid y, x; \theta\left(\cdot\right), \delta\left(\cdot\right)\right) = \frac{\prod_{t=1}^T f_{Y_t|X_t,\eta}\left(y_t \mid x_t, \eta; \theta\left(\cdot\right)\right) f_{\eta|X}\left(\eta \mid x; \delta\left(\cdot\right)\right)}{\int \prod_{t=1}^T f_{Y_t|X_t,\eta}\left(y_t \mid x_t, \widetilde{\eta}; \theta\left(\cdot\right)\right) f_{\eta|X}\left(\widetilde{\eta} \mid x; \delta\left(\cdot\right)\right) d\widetilde{\eta}}, \tag{17}$$

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

Let $\psi_\tau(u) = \tau - \mathbf{1}\{u < 0\}$. The function 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):

$$\rho_\tau(u) = (\tau - \mathbf{1}\{u < 0\})\, u, \qquad \psi_\tau(u) = \frac{d\rho_\tau(u)}{du}.$$

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:

$$\mathbb{E}\left[\sum_{t=1}^{T} W_{it}(\eta_i)\, \psi_\tau\left(Y_{it} - W_{it}(\eta_i)'\theta(\tau)\right)\right] = 0, \tag{18}$$

and:

$$\mathbb{E}\left[Z_i \psi_\tau\left(\eta_i - Z_i'\delta(\tau)\right)\right] = 0. \tag{19}$$

Indeed, (18) is the first-order condition associated with the infeasible population quantile regression of Y_{it} on $W_{it}(\eta_i)$. Similarly, (19) corresponds to the infeasible quantile regression of η_i on Z_i.

Applying the law of iterated expectations to (18) and (19), respectively, we obtain the following integrated moment restrictions, for all $\tau \in (0,1)$:

$$\mathbb{E}\left[\int \left(\sum_{t=1}^{T} W_{it}(\eta)\, \psi_\tau\left(Y_{it} - W_{it}(\eta)'\theta(\tau)\right)\right) f\left(\eta \mid Y_i, X_i; \theta(\cdot), \delta(\cdot)\right) d\eta\right] = 0, \tag{20}$$

and:

$$\mathbb{E}\left[\int \left(Z_i \psi_\tau\left(\eta - Z_i'\delta(\tau)\right)\right) f\left(\eta \mid Y_i, X_i; \theta(\cdot), \delta(\cdot)\right) d\eta\right] = 0, \tag{21}$$

where, here and in the rest of the analysis, we use f as a shorthand for the posterior density $f_{\eta|Y,X}$.

It follows from (20)-(21) that, if the posterior density of the individual effects were known, then estimating the model's parameters could be done using two sets of linear quantile regressions, weighted by the posterior density. However, as the notation makes clear, the posterior density in (17) depends on the entire processes $\theta(\cdot)$ and $\delta(\cdot)$. Specifically we have, for absolutely continuous conditional densities of outcomes and individual effects:

$$f_{Y_t|X_t,\eta}\left(y_t \mid x_t, \eta; \theta(\cdot)\right) = \lim_{\epsilon \to 0} \frac{\epsilon}{w_t(\eta)'\left[\theta(u_t + \epsilon) - \theta(u_t)\right]}, \tag{22}$$

and:

$$f_{\eta|X}\left(\eta \mid x;\delta\left(\cdot\right)\right)=\lim_{\epsilon\rightarrow 0}\frac{\epsilon}{z'\left[\delta\left(v+\epsilon\right)-\delta\left(v\right)\right]}, \tag{23}$$

where u_t and v are defined by $w_t\left(\eta\right)'\theta\left(u_t\right)=y_t$ and $z'\delta\left(v\right)=\eta$, respectively. Equations (22) and (23) 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 directly recover $\theta\left(\tau\right)$ and $\delta\left(\tau\right)$ in (20)-(21) in a τ-by-τ fashion. The main idea of the algorithm that we present in the next subsection 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, an 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.

Parametric specification. Building on Wei and Carroll (2009), we approximate $\theta\left(\cdot\right)$ and $\delta\left(\cdot\right)$ 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 $\theta\left(\tau\right)$ and $\delta\left(\tau\right)$ as functions of $\tau\in(0,1)$. For fixed L, the spline specification may be seen as an approximation to the underlying quantile functions.

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

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

The approximating model depends on the finite-dimensional parameter vector ξ that is used to construct interpolating splines. The associated likelihood function and density of individual effects are then denoted as $f_{Y_t|X_t,\eta}\left(y_t \mid x_t,\eta;\xi_A\right)$ and $f_{\eta|X}\left(\eta \mid x;\xi_B\right)$, respectively, and the implied posterior density is:

$$f\left(\eta \mid y,x;\xi\right)=\frac{\prod_{t=1}^{T}f_{Y_t|X_t,\eta}\left(y_t \mid x_t,\eta;\xi_A\right)f_{\eta|X}\left(\eta \mid x;\xi_B\right)}{\int\prod_{t=1}^{T}f_{Y_t|X_t,\eta}\left(y_t \mid x_t,\widetilde{\eta};\xi_A\right)f_{\eta|X}\left(\widetilde{\eta} \mid x;\xi_B\right)d\widetilde{\eta}}. \tag{24}$$

The approximating densities take closed-form expressions when using piecewise-linear splines. Moreover, when implementing the algorithm in practice we augment the specification with parametric models in the tail intervals of the coefficients of $\theta(\tau)$ and $\delta(\tau)$ corresponding to the constant terms. In this case the estimation algorithm needs to be modified slightly. See Section 6.1 for a discussion of implementation.

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

$$\mathbb{E}\left[\int \left(\sum_{t=1}^{T} W_{it}(\eta)\, \psi_{\tau_\ell}\left(Y_{it} - W_{it}(\eta)' \theta(\tau_\ell)\right)\right) f(\eta \mid Y_i, X_i; \xi)\, d\eta\right] = 0, \tag{25}$$

and:

$$\mathbb{E}\left[\int \left(Z_i \psi_{\tau_\ell}(\eta - Z_i'\delta(\tau_\ell))\right) f(\eta \mid Y_i, X_i; \xi)\, d\eta\right] \; = \; 0. \tag{26}$$

3.2 Estimation algorithm

Let (Y_i, X_i'), $i = 1, ..., N$, be an i.i.d. sample. Motivated by the integrated moment restrictions (25)-(26) we propose to estimate the model's parameters by using an iterative method. In practice we use a simulation-based approach to replace the integrals in (25)-(26) by sums. Starting with initial parameter values $\widehat{\xi}^{(0)}$, we iterate the following two steps until convergence to a stationary distribution.

Algorithm. *(stochastic EM)*

1. *For all $i = 1, ..., N$, compute the posterior density:*

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

and draw M values $\eta_i^{(1)}$,...,$\eta_i^{(M)}$ from $\widehat{f}_i^{(s)}$.

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

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

This sequential simulated method-of-moment method is related to, but different from, the standard EM algorithm (Dempster *et al.*, 1977). As 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 second step of the algorithm (the "M"-step) 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 M-step problem nicely decomposes into L different τ_ℓ-specific subproblems. Secondly, using the check function yields a globally convex objective function in each step. In fact, the "M"-step simply consists of $2L$ ordinary quantile regressions, where the simulated values of the individual effects are treated, in turn, as covariates and dependent variables.

At the same time, 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, even as the number of draws M tends to infinity.[8]

Second, unlike in deterministic versions of EM, in the "E"-step we draw M values for the individual effects according to their posterior density $\widehat{f}_i^{(s)}(\eta) = f\left(\eta \mid Y_i, X_i; \widehat{\xi}^{(s)}\right)$. We use a random-walk Metropolis-Hastings sampler for this purpose, but other choices are possible (such as particle filter methods).[9] 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. In practice, we stop the chain after a large number of iterations and we report an average across the last $\widetilde{S}$ values: $\widehat{\xi} = \frac{1}{\widetilde{S}} \sum_{s=S-\widetilde{S}+1}^{S} \widehat{\xi}^{(s)}$.

In each iteration of the algorithm, the draws $\eta_i^{(1)}$,...,$\eta_i^{(M)}$ are randomly re-drawn. This approach, sometimes referred to as "stochastic EM", thus differs from the simulated EM algorithm of McFadden and Ruud (1994) where the same underlying uniform draws are used in each iteration. Nielsen (2000a, 2000b) studies and compares various statistical properties of simulated EM and stochastic EM in a likelihood context. In particular, he provides conditions under which the Markov chain output of stochastic EM is ergodic. 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). Building on Nielsen's work, we next analyze the statistical

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

[9]Note that the posterior density is non-negative by construction. In particular, drawing from $\widehat{f}_i^{(s)}(\eta)$ automatically produces rearrangement of the various quantile curves, as in Chernozhukov, Galichon and Fernández-Val (2010).

properties of estimators based on fixed-M and large-M versions of the algorithm.

3.3 Asymptotic properties

We now discuss the asymptotic properties of the estimation algorithm. Throughout, T is fixed while N tends to infinity.

Parametric inference. We start by discussing the asymptotic properties of the estimator based on the stochastic EM algorithm, for fixed number of draws M, in the case where the parametric model is assumed to be correctly specified. That is, K_1, K_2 (the number of series terms) and L (the size of the grid on the unit interval) are held fixed as N tends to infinity. In the next paragraph we will study consistency as K_1, K_2 and L tend to infinity with N, in the large-M limit.

Nielsen (2000a) studies the statistical properties of the stochastic EM algorithm in a likelihood case. He provides conditions under which the Markov Chain $\widehat{\xi}^{(s)}$ is ergodic, for a fixed sample size. In addition, he also characterizes the asymptotic distribution of $\sqrt{N}\left(\widehat{\xi}^{(s)} - \overline{\xi}\right)$ as N increases, where $\overline{\xi}$ denotes the population parameter vector.

In Appendix B we rely on Nielsen's work to characterize the asymptotic distribution of $\widehat{\xi}^{(s)} = ((\widehat{\theta}^{(s)})', (\widehat{\delta}^{(s)})')'$ in our model, where the optimization step is not likelihood-based but relies on quantile-based estimating equations. Specifically, if s corresponds to a draw from the ergodic distribution of the Markov Chain, and M is the number of draws per iteration, then:

$$\sqrt{N}\left(\widehat{\xi}^{(s)} - \overline{\xi}\right) \stackrel{d}{\to} \mathcal{N}(0, \mathcal{V} + \mathcal{V}_M),$$

where the expressions of $\mathcal{V}$ and $\mathcal{V}_M$ are given in Appendix B.

In addition, if $\widehat{\xi}$ is a parameter draw and M tends to infinity, or alternatively if $\widehat{\xi}$ is computed as the average of $\widehat{\xi}^{(s)}$ over $\widetilde{S}$ iterations with $\widetilde{S}$ tending to infinity (as in our implementation), then:

$$\sqrt{N}\left(\widehat{\xi} - \overline{\xi}\right) \stackrel{d}{\to} \mathcal{N}(0, \mathcal{V}),$$

$\mathcal{V}$ being the asymptotic variance of the method-of-moments estimator based on the integrated moment restrictions (25)-(26).

Nonparametric consistency. In the asymptotic theory of the previous paragraph, K_1, K_2 and L are held fixed as N tends to infinity. It may be more appealing to see the parametric specification based on series and splines as an approximation to the quantile functions, which becomes more accurate as the dimensions K_1, K_2 and L increase. Here our aim is to provide conditions under which the estimator is consistent as N, K_1, K_2, and L tend to infinity.

To proceed we consider the following assumption on the data generating process, as in Belloni, Chernozhukov and Fernández-Val (2011):

$$Y_{it} = W_{it}(\eta_i)'\overline{\theta}(U_{it}) + R_Y(X_{it}, \eta_i, U_{it}),$$

and, similarly:

$$\eta_i = Z_i'\overline{\delta}(V_i) + R_\eta(X_i, V_i),$$

where $\sup_{(x,e,u)} |R_Y(x, e, u)| = o(1)$ as K_1 tends to infinity, and $\sup_{(x,v)} |R_\eta(x, v)| = o(1)$ as K_2 tends to infinity.

Let $\xi(\tau) = (\theta(\tau)', \delta(\tau)')'$ be a $(K_1 + K_2) \times 1$ vector for all $\tau \in (0, 1)$, and let $\xi : (0, 1) \to \mathbb{R}^{K_1+K_2}$ be the associated function. Let us consider the estimator $\widehat{\xi} = (\widehat{\theta}', \widehat{\delta}')'$ based on the integrated moment restrictions (25)-(26). This analysis as $M \to \infty$ thus ignores the impact of small-M simulation error. Note that $\widehat{\xi}$ is a function defined on the unit interval. In Appendix B we provide and discuss conditions that guarantee that $\widehat{\xi}$ is uniformly consistent for $\overline{\xi} = (\overline{\theta}', \overline{\delta}')'$, that is:

$$\sup_{\tau\in(0,1)} \left\| \widehat{\xi}(\tau) - \overline{\xi}(\tau) \right\| = o_p(1), \tag{28}$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^{K_1+K_2}$.

Some of the conditions for consistency given in Appendix B are non-primitive. In particular, an identification condition is required which is related to Assumption 3, though it differs from it due to the fact that our estimator is based on a set of moment conditions rather than the likelihood. More generally, models with latent distributions such as the nonlinear panel data models we analyze in this paper are subject to ill-posedness, making a complete characterization of asymptotic distributions challenging.[10] A practical possibility, for which we do not yet have a formal justification, is to use empirical counterparts of the fixed-(K_1, K_2, L) asymptotic formulas derived in the previous paragraph, or alternatively the bootstrap, to conduct inference. A related question is that of the practical choice of K_1, K_2 and L. In this paper we do not characterize the asymptotic distribution of our estimator as N, K_1, K_2, and L tend to infinity, and we leave these important questions to future work.

[10] In particular, the class of models we consider nests nonparametric deconvolution models with repeated measurements (Kotlarski, 1967, Horowitz and Markatou, 1996, Delaigle, Hall and Meister, 2008, Bonhomme and Robin, 2010). In such settings, quantiles are generally not root-N estimable (Hall and Lahiri, 2008).

4 Dynamic models

In this section we extend the method to dynamic models with dependence on lagged outcomes or predetermined covariates.

4.1 Models, examples, and identification

In a dynamic extension of the static model (3), we specify the conditional quantile function of Y_{it} given $Y_{i,t-1}$, X_{it} and η_i as:

$$Y_{it} = Q_Y\left(Y_{i,t-1}, X_{it}, \eta_i, U_{it}\right), \quad i = 1, ..., N, \quad t = 2, ..., T. \tag{29}$$

A simple extension is obtained by replacing $Y_{i,t-1}$ by a vector containing various lags of the outcome variable. As in the static case, Q_Y could depend on t.

Linear versions of (29) are widely used in applications, including in the study of individual earnings, firm-level investment, cross-country growth, or in the numerous applications of panel VAR models. In these applications, interactions between heterogeneity and dynamics are often of great interest. A recent example is the analysis of institutions and economic growth in Acemoglu *et al.* (2015).

The assumptions we impose in model (29), and the modelling of unobserved heterogeneity, both depend on the nature of the covariates process. We consider two cases in turn: strictly exogenous and predetermined covariates.

Autoregressive models. In the case where covariates are strictly exogenous, with some abuse of notation we suppose that Assumption 1 holds with $(Y_{i,t-1}, X_{it}')'$ instead of X_{it} and $(Y_{i1}, X_{i1}', ..., X_{iT}')'$ instead of X_i. Note that the latter contains both strictly exogenous covariates and first-period outcomes. Individual effects can be written without loss of generality as:

$$\eta_i = Q_\eta\left(Y_{i1}, X_i, V_i\right), \quad i = 1, ..., N, \tag{30}$$

and we suppose that Assumption 2 holds with $(Y_{i1}, X_i')'$ instead of X_i.

Predetermined covariates. In dynamic models with predetermined regressors, current values of U_{it} may affect future values of covariates X_{is}, $s > t$. Given the presence of latent variables in our nonlinear setup, a model for the feedback process is needed. That is, we need to specify the conditional distribution of X_{it} given $(Y_i^{t-1}, X_i^{t-1}, \eta_i)$, where $Y_i^{t-1} =$

$(Y_{i,t-1},...,Y_{i1})'$ and $X_i^{t-1} = (X'_{i,t-1},...,X'_{i1})'$. We use additional quantile specifications for this purpose. In the case where X_{it} is scalar, and under a conditional first-order Markov assumption for (Y_{it}, X_{it}), $t = 1,...,T$, given η_i, we specify, without further loss of generality:

$$X_{it} = Q_X\left(Y_{i,t-1}, X_{i,t-1}, \eta_i, A_{it}\right), \quad i = 1,...,N, \quad t = 2,...,T. \tag{31}$$

We suppose that Assumptions 1 and 2 hold, with $(Y_{i,t-1}, X'_{it})'$ instead of X_{it} and $(Y_{i1}, X'_{i1})'$ instead of X_i, and:

$$\eta_i = Q_\eta\left(Y_{i1}, X_{i1}, V_i\right), \quad i = 1,...,N. \tag{32}$$

We then complete the model with the following assumption on the feedback process.

Assumption 4. *(predetermined covariates)*

(i) A_{it} follows a standard uniform distribution, independent of $(Y_{i,t-1}, X_{i,t-1}, \eta_i)$.

(ii) $\tau \mapsto Q_X(y, x, \eta, \tau)$ is strictly increasing on $(0,1)$, for almost all (y, x, η) in the support of $(Y_{i,t-1}, X_{i,t-1}, \eta_i)$.

(iii) For all $t \neq s$, A_{it} is independent of A_{is}.

Model (31) can be extended to multi-dimensional predetermined covariates using a triangular approach in the spirit of the one introduced in Example 3. For example, with two-dimensional $X_{it} = (X_{1it}, X_{2it})'$:

$$\begin{aligned} X_{1it} &= Q_{X_1}\left(Y_{i,t-1}, X_{1i,t-1}, X_{2i,t-1}, \eta_i, A_{1it}\right), \\ X_{2it} &= Q_{X_2}\left(Y_{i,t-1}, X_{1it}, X_{1i,t-1}, X_{2i,t-1}, \eta_i, A_{2it}\right), \end{aligned} \tag{33}$$

where η_i may be scalar or multi-dimensional as in Example 3.

Example 4: Panel quantile autoregression. A dynamic counterpart to Example 2 is the following linear quantile regression model:

$$Y_{it} = \rho\left(U_{it}\right) Y_{i,t-1} + X'_{it}\beta\left(U_{it}\right) + \eta_i\gamma\left(U_{it}\right). \tag{34}$$

Model (34) differs from the more general model studied in Galvao (2011):

$$Y_{it} = \rho\left(U_{it}\right) Y_{i,t-1} + X'_{it}\beta\left(U_{it}\right) + \eta_i\left(U_{it}\right). \tag{35}$$

Similarly as in (7), and in contrast with the models introduced in this paper, the presence of the functional heterogeneity term $\eta_i(\tau)$ makes fixed-T consistent estimation problematic in (35).

An extension of (34) is:

$$Y_{it} = h\left(Y_{i,t-1}\right)'\rho\left(U_{it}\right) + X_{it}'\beta\left(U_{it}\right) + \eta_i\gamma\left(U_{it}\right), \quad t = 2, ..., T, \tag{36}$$

where h is a univariate function. For example, when $h(y) = |y|$ model (36) is a panel data version of the CAViaR model of Engle and Manganelli (2004). Other choices will lead to panel counterparts of various dynamic quantile models (e.g., Gouriéroux and Jasiak, 2008). The approach developed in this paper allows for more general, nonlinear series specifications of dynamic quantile functions in a panel data context.

Example 5: Quantile autoregression with predetermined covariates. Extending Example 4 to allow for a scalar predetermined covariate X_{it}, we may augment (34) with the following linear quantile specification for X_{it}:

$$X_{it} = \mu\left(A_{it}\right)Y_{i,t-1} + \xi_1\left(A_{it}\right)X_{i,t-1} + \xi_0\left(A_{it}\right) + \zeta\left(A_{it}\right)\eta_i.$$

This specification can be extended to allow for multi-dimensional predetermined regressors, as in (33).

Identification. In dynamic models 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:

$$\begin{aligned} f_{Y_1,Y_2,Y_3,Y_4|X}\left(y_1, y_2, y_3, y_4 \mid x\right) &= \int f_{Y_2|Y_1,\eta,X}\left(y_2 \mid y_1, \eta, x\right) f_{Y_3|Y_2,\eta,X}\left(y_3 \mid y_2, \eta, x\right) \\ &\quad \times f_{Y_4|Y_3,\eta,X}\left(y_4 \mid y_3, \eta, x\right) f_{\eta,Y_1|X}\left(\eta, y_1 \mid x\right) d\eta, \end{aligned} \tag{37}$$

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 A.[11]

[11] In the dynamic model (36), it follows from Hu and Shum (2012)'s analysis that one can rely on (14) as in the static case, provided the averages across τ values of the coefficients of exogenous regressors and lagged

Lastly, autoregressive models with predetermined covariates can be shown to be nonparametrically identified using similar arguments, provided the feedback process is first-order Markov.

4.2 Estimation in dynamic models

The estimation algorithm of Section 3 can be directly modified to deal with autoregressive models with strictly exogenous covariates. Consider a linear specification of the quantile functions (29) and (30), possibly based on series. Then the stochastic EM algorithm essentially takes the same form as in the static case, except for the posterior density of the individual effects which is now computed as:

$$f(\eta \mid y, x; \xi) = \frac{\prod_{t=2}^{T} f_{Y_t|Y_{t-1},X_t,\eta}(y_t \mid y_{t-1}, x_t, \eta; \xi_A)\, f_{\eta|Y_1,X}(\eta \mid y_1, x; \xi_B)}{\int \prod_{t=2}^{T} f_{Y_t|Y_{t-1},X_t,\eta}(y_t \mid y_{t-1}, x_t, \widetilde{\eta}; \xi_A)\, f_{\eta|Y_1,X}(\widetilde{\eta} \mid y_1, x; \xi_B)\, d\widetilde{\eta}}. \tag{38}$$

General predetermined regressors. In models with predetermined covariates, the critical difference is in the nature of the posterior density of the individual effects. Letting $W_{it} = (Y_{it}, X_{it}')'$ and $W_i = (W_{i1}', ..., W_{iT}')'$ we have:

$$\begin{aligned} f(\eta \mid y, x; \xi) &= \frac{f_{W_2,...,W_T}(w_2, ..., w_T \mid w_1, \eta)\, f_{\eta|W_1}(\eta \mid w_1)}{\displaystyle\int f_{W_2,...,W_T}(w_2, ..., w_T \mid w_1, \eta)\, f_{\eta|W_1}(\eta \mid w_1)\, d\eta} \\ &= \frac{f_{\eta|W_1}(\eta \mid w_1; \xi_B) \prod_{t=2}^{T} f_{Y_t|Y_{t-1},X_t,\eta}(y_t \mid y_{t-1}, x_t, \eta; \xi_A)\, f_{X_t|W^{t-1},\eta}(x_t \mid w^{t-1}, \eta; \xi_C)}{\displaystyle\int f_{\eta|W_1}(\widetilde{\eta} \mid w_1; \xi_B) \prod_{t=2}^{T} f_{Y_t|Y_{t-1},X_t,\eta}(y_t \mid y_{t-1}, x_t, \widetilde{\eta}; \xi_A)\, f_{X_t|W^{t-1},\eta}(x_t \mid w^{t-1}, \widetilde{\eta}; \xi_C)\, d\widetilde{\eta}}, \end{aligned}$$

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 quantile model only specifies the partial likelihood:

$$\prod_{t=2}^{T} f_{Y_t|Y_{t-1},X_t,\eta}(y_t \mid y_{t-1}, x_t, \eta; \xi_A).$$

However, the posterior density of the individual effects also depends on the feedback process:

$$f_{X_t|W^{t-1},\eta}\left(x_t \mid w^{t-1}, \eta; \xi_C\right),$$

outcome are identified based on:

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

in addition to the density of individual effects. Note that the feedback process could depend on an additional vector of individual effects different from η_i.

In line with our approach, we also specify the quantile function of covariates in (31) using linear (series) quantile regression models. Specifically, letting X_{pit}, $p = 1, ..., P$, denote the various components of X_{it}, we specify the following triangular, recursive system that extends Example 5 to multi-dimensional predetermined covariates:

$$\begin{aligned} X_{1it} &= W_{1it}(\eta_i)\mu_1\left(A_{1it}\right), \\ \ldots \quad &\ldots \quad \ldots \\ X_{Pit} &= W_{Pit}(\eta_i)\mu_P\left(A_{Pit}\right), \end{aligned} \tag{39}$$

where $A_{1it}, ..., A_{Pit}$ follow independent standard uniform distributions, independent of all other random variables in the model, $W_{1it}(\eta_i)$ contains functions of $(Y_{i,t-1}, X_{i,t-1}, \eta_i)$, and $W_{pit}(\eta_i)$ contains functions of $(X_{1it}, ...X_{p-1,it}, Y_{i,t-1}, X_{i,t-1}, \eta_i)$ for $p > 1$. The parameter vector ξ_C includes all $\mu_p(\tau_\ell)$, for $p = 1, ..., P$ and $\ell = 1, ..., L$.

The model with predetermined regressors has thus three layers of quantile regressions: the outcome model (29) specified as a linear quantile regression, the model of the feedback process (39), and the model of individual effects (32), which here depends on first-period outcomes and covariates. The estimation algorithm is similar to the one for static models, with minor differences in both steps.[12]

5 Quantile marginal effects

In nonlinear panel data models, it is often of interest to compute the effect of marginal changes in covariates on the entire distribution of outcome variables. As an example, let us consider the following average quantile marginal effect (QME hereafter) for continuous X_{it}:

$$M(\tau) = \mathbb{E}\left[\frac{\partial Q_Y(X_{it}, \eta_i, \tau)}{\partial x}\right],$$

where $\partial Q_Y/\partial x$ denotes the vector of partial derivatives of Q_Y with respect to its first $\dim(X_{it})$ arguments.

In the quantile regression model of Example 2, individual quantile marginal effects are equal to $\partial Q_Y(X_{it}, \eta_i, \tau)/\partial x = \beta(\tau)$, and $M(\tau) = \beta(\tau)$. In Example 3, individual QME are

[12] In addition, in Appendix C we describe how to allow for autocorrelated errors in model (3)-(4).

heterogeneous, equal to $\beta(\tau)+\gamma_2(\tau)\eta_{i2}$, and $M(\tau) = \beta(\tau)+\gamma_2(\tau)\mathbb{E}\left[\eta_{i2}\right]$. Series specifications of the quantile function as in (1) can allow for rich heterogeneity in individual QME.

Dynamic models. Quantile marginal effects are also of interest in dynamic models. One can define short-run average QME as:

$$M_t(\tau) = \mathbb{E}\left[\frac{\partial Q_Y(Y_{i,t-1}, X_{it}, \eta_i, \tau)}{\partial x}\right].$$

Moreover, when considering marginal changes in the lagged outcome $Y_{i,t-1}$, the average QME, $\mathbb{E}\left[\partial Q_Y(Y_{i,t-1}, X_{it}, \eta_i, \tau)/\partial y\right]$, can be interpreted as a nonlinear measure of state dependence. In that case $\partial Q_Y/\partial y$ denotes the derivative of Q_Y with respect to its first argument.

Dynamic models also provide the opportunity to document dynamic quantile marginal effects, such as the following one-period-ahead average QME:

$$M_{t+1/t}(\tau_1,\tau_2) = \mathbb{E}\left[\frac{\partial Q_Y(Q_Y(Y_{i,t-1}, X_{it}, \eta_i, \tau_1), X_{i,t+1}, \eta_i, \tau_2)}{\partial y} \times \frac{\partial Q_Y(Y_{i,t-1}, X_{it}, \eta_i, \tau_1)}{\partial x}\right].$$

$M_{t+1/t}(\tau_1,\tau_2)$ measures the average effect of a marginal change in X_{it} when η_i is kept fixed, and the innovations in periods t and $t+1$ have rank τ_1 and τ_2, respectively.

Panel quantile treatment effects. When the covariate of interest is binary, as in our empirical application in Section 6, one can define panel data versions of quantile treatment effects. To see this, let D_{it} be the binary covariate of interest, and let X_{it} include all other time-varying covariates. Consider the static model (3), the argument extending directly to dynamic models. Potential outcomes are defined as:

$$Y_{it}(d) = Q_Y\left(d, X_{it}, \eta_i, U_{it}\right), \quad d \in \{0,1\}.$$

Under Assumption 1, $(Y_{it}(0), Y_{it}(1))$ is conditionally independent of D_{it} given (X_i, η_i). This amounts to assuming selection on observables and unobservables, when unobserved effects η_i are identified from the panel dimension.

The average conditional quantile treatment effect is then defined as:

$$\mathbb{E}\left[Q_Y\left(1, X_{it}, \eta_i, \tau\right) - Q_Y\left(0, X_{it}, \eta_i, \tau\right)\right].$$

In the linear quantile regression model of Example 2, this is simply the coefficient of the vector $\beta(\tau)$ corresponding to D_{it}. In fact, the distribution of treatment effects is identified for this model, under the conditions spelled out in Section 2. The key assumption is rank invariance of U_{it} given X_i and η_i.

It is also possible to define unconditional quantile treatment effects, as:

$$F_{Y_{it}(1)}^{-1}(\tau) - F_{Y_{it}(0)}^{-1}(\tau),$$

where the cdfs $F_{Y_{it}(0)}$ and $F_{Y_{it}(1)}$ are given by:[13]

$$F_{Y_{it}(d)}(y) = \mathbb{E}\left[\int_0^1 \mathbf{1}\left\{Q_Y(d, X_{it}, \eta_i, \tau) \leq y\right\} d\tau\right], \quad d \in \{0,1\}. \tag{40}$$

All these quantities can readily be estimated using our panel quantile estimator.

6 Empirical application

In this section we present an empirical illustration to the link between mothers' smoking during pregnancy and birthweight. We start by discussing how we implement the estimation algorithm in practice.

6.1 Implementation

Piecewise-linear splines. We use piecewise-linear splines as an approximating model. Although other spline families could be used instead, computing the implied likelihood functions would then require inverting quantile functions numerically. In contrast, for linear splines we have, for all $\ell = 1, ..., L-1$:

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

and the implied approximating period-t density of outcomes and the implied approximating density of individual effects take simple closed-form expressions:

$$f_{Y_t|X_t,\eta}(y_t \mid x_t, \eta; \xi_A) = \frac{\tau_{\ell+1} - \tau_\ell}{w_t(\eta)'\left[\theta(\tau_{\ell+1}) - \theta(\tau_\ell)\right]} \quad \text{if } w_t(\eta)'\theta(\tau_\ell) < y_t \leq w_t(\eta)'\theta(\tau_{\ell+1}), \tag{41}$$

$$f_{\eta|X}(\eta \mid x; \xi_B) = \frac{\tau_{\ell+1} - \tau_\ell}{z'\left[\delta(\tau_{\ell+1}) - \delta(\tau_\ell)\right]} \quad \text{if } z'\delta(\tau_\ell) < \eta \leq z'\delta(\tau_{\ell+1}), \tag{42}$$

augmented with a specification in the tail intervals $(0, \tau_1)$ and $(\tau_L, 1)$.

[13]Note that unconditional quantile treatment effects cannot be directly estimated as in Firpo (2007) in this context, due to the presence of the unobserved η_i and the lack of fixed-T identification for fixed-effects binary choice models.

Tail intervals. In order to model quantile functions in the intervals $(0, \tau_1)$ and $(\tau_L, 1)$ one could 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 D we outline a different, exponential-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 empirical application.

6.2 Application: smoking and birthweight

Here 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 nonlinearities in the relationship between smoking and weight at birth. As a complement, in Appendix E we report the results of a Monte Carlo simulation broadly calibrated to this application, in order to assess the performance of our estimator in finite samples.

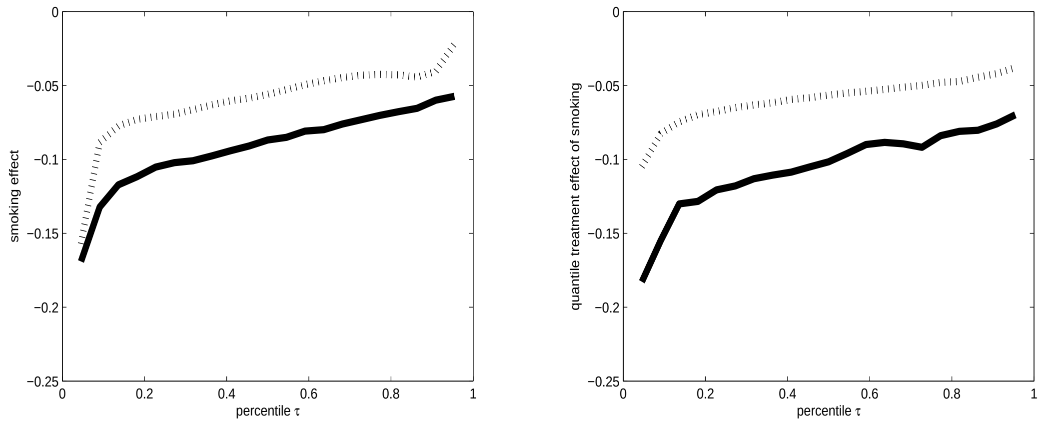
We focus on a balanced subsample from the US natality data used in Abrevaya (2006), which comprises 12360 women with 3 children each. Our outcome is the log-birthweight. The main covariate is a binary smoking indicator. Age of the mother and gender of the child are used as additional controls.

An OLS regression yields a significantly 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 the results in 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 left graph of Figure 1 shows the quantile estimate of the smoking effect. We use a linear quantile regression specification as in Example 2, augmented with a parametric exponential model in the tail intervals. The covariates are smoking status, age, and gender, with an intercept. We use individual-specific averages of these variables as covariates in the specification for η_i. Estimates are computed using $L = 21$ knots. The stochastic EM algorithm is run for 100 iterations, with 500 random walk Metropolis-Hastings

Figure 1: Quantile effects of smoking during pregnancy on log-birthweight (linear quantile specification)



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

draws within each iteration.[14] Parameter estimates are computed as averages of the 50 last iterations of the algorithm.[15]

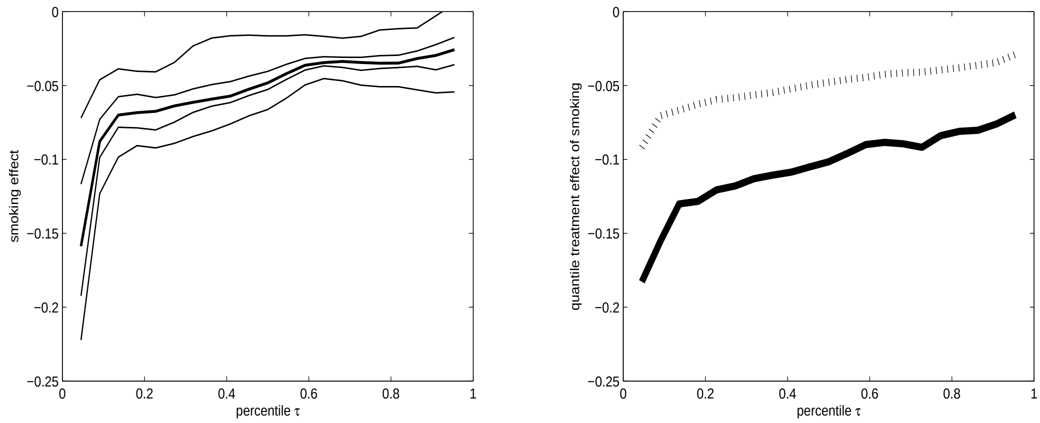
We see on the left graph of Figure 1 that the smoking effect becomes less negative when correcting for time-invariant endogeneity through the introduction of mother-specific fixed-effects. At the same time, the effect is still sizable, and it remains increasing along the distribution.

As another exercise, on the right graph of Figure 1 we compute the unconditional 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 (that is, observed X_i and unobserved η_i) constant, as defined in Section 5. We report differences in quantiles of simulated potential outcomes obtained using the method of Machado and Mata (2005). This exercise illustrates the usefulness of specifying and estimating a complete semiparametric model of the joint distribution of outcomes and unobservables, in order to

[14] The variance of the random walk proposal is set to achieve $\approx 30\%$ acceptance rate.

[15] For θ's, starting parameter values are taken based on ordinary quantile regressions of log-birthweight on smoking status, age, and gender, with an intercept, setting the coefficient of η_i in the outcome equation to one. For δ's, we set all initial quantile parameters to $\{.1, .2, ..., 2.1\}$. The initial values for the exponential parameters in the tails are all set to 20. We experimented with other starting values for the model's parameters (for example, we initialized the δ's based on quantile regressions of individual-specific means $\overline{Y}_i$ on $\overline{X}_i$) and found no qualitative differences compared to the results we report.

Figure 2: Quantile effects of smoking during pregnancy on log-birthweight (interacted quantile specification)



Note: Data from Abrevaya (2006). Left graph: lines represent the percentiles .05, .25, .50, .75, and .95 of the heterogeneous smoking effect across mothers, at various percentiles τ. Right graph: solid line is the raw quantile treatment effect of smoking; dashed line is the quantile treatment effect estimate based on panel quantile regression with interactions.

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 on the right graph of Figure 1 are broadly similar to the ones reported on the left graph. An interesting finding is that in this case the endogeneity bias (that is, the difference between the dashed and solid lines) is slightly larger, and that it tends to decrease as one moves from lower to higher quantiles of birthweight.

Lastly, on Figure 2 we report the results of an interacted quantile model, as in (1) and (2), where the specification allows for all first-order interactions between covariates (that is, smoking status, age and gender) and the unobserved mother-specific effect. In this model the quantile effect of smoking is mother-specific. The results on the right graph show the unconditional quantile treatment effect of smoking. Results are similar to the ones obtained for a simple linear specification (see the right graph of Figure 1). However, on the left graph of Figure 2 we see substantial mother-specific heterogeneity in the conditional quantile treatment effect of smoking, as for some mothers smoking appears particularly detrimental to children's birthweight, whereas for other mothers the smoking effect, while consistently negative, is much smaller. This evidence is in line with the results of a linear random coefficients model reported in Arellano and Bonhomme (2012).

7 Conclusion

Quantile methods are flexible tools to model nonlinear panel data relationships. In this work, quantile regression is used to model the dependence between outcomes, covariates and individual heterogeneity, and between individual effects and exogenous regressors or initial conditions. Quantile specifications also allow modelling 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.

Our approach leads to fixed-T identification of complete models. The estimation algorithm exploits the computational advantages of linear quantile regression, within an iterative scheme which allows dealing with the presence of unobserved individual effects. Beyond static or dynamic quantile regression models with single or multiple individual effects, our approach naturally extends to series specifications, thus allowing for rich interactions between covariates and heterogeneity at various points of the distribution.

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:

$$f_{Y|X}(y|x;\alpha,\delta(\cdot)) = \int f_{Y|X,\eta}(y|x,\eta;\alpha) f_{\eta|X}(\eta|x;\delta(\cdot))d\eta,$$

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

Our framework also 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. Arellano, Blundell and Bonhomme (2015) use a quantile-based approach to document nonlinear relationships between earnings shocks to households and their lifetime profiles of earnings and consumption. This application illustrates the potential of our estimation approach in dynamic settings.

A relevant issue for empirical practice is measurement error. Our approach may be extended to allow covariates to be measured with error, as the analysis in Wei and Carroll (2009) illustrates. When a validation sample is available, our algorithm can also be modified to allow for measurement error in outcome variables. In both cases, true variables are treated similarly as latent individual effects in the above analysis, and they are repeatedly drawn from their posterior densities in each iteration of the algorithm.

Lastly, this paper leaves a number of important questions unanswered. Statistical inference in the nonparametric problem, where the complexity of the approximating model increases together with the sample size, is one of them. Providing primitive conditions for identification, and devising efficient computational routines, are other important questions for future work.

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] Acemoglu, D., S. Naidu, P. Restrepo, and J. Robinson (2015): "Democracy Does Cause Growth," unpublished manuscript.

[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,

[7] Arellano, M., R. Blundell, and S. Bonhomme (2015): "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework," unpublished manuscript.

[8] Arellano, M. and S. Bonhomme (2011): "Nonlinear Panel Data Analysis," *Annual Review of Economics*, 3, 2011, 395-424. 1823–1868.

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

[10] Arellano, M., and M. Weidner (2015): "Instrumental Variable Quantile Regressions in Large Panels with Fixed Effects," unpublished manuscript.

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

[12] 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.

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

[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] Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics*, 6, 5559–5632.

[18] Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010): "Quantile and Probability Curves without Crossing ," *Econometrica*, 78(3), 1093–1125.

[19] Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81(2), 535–580.

[20] Chernozhukov, V., I. Fernández-Val, S. Hoderlein, H. Holzmann, and W. Newey (2015): "Nonparametric Identification in Panels using Quantiles," *Journal of Econometrics*, forthcoming.

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

[22] 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.

[23] D'Haultfoeuille, X. (2011): "On the Completeness Condition for Nonparametric Instrumental Problems," *Econometric Theory*, 27, 460–471

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

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

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

[27] Firpo, S. (2007): “Efficient Semiparametric Estimation of Quantile Treatment Effects,” *Econometrica*, 75(1), 259–276.

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

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

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

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

[32] Graham, B., J. Hahn, A. Poirier, and J. Powell (2015): “Quantile Regression with Panel Data,” unpublished manuscript.

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

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

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

[36] Hu, Y. (2015): “Microeconomic Models with Latent Variables: Applications of Measurement Error Models in Empirical Industrial Organization and Labor Economics,” Technical report, Cemmap Working Papers, CWP03/15.

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

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

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

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

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

[42] Koenker, R. (2005): *Quantile Regression*, Monograph of the Econometric Society.

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

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

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

[46] Machado, J. A. F., and J. Mata (2005): "Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression," *Journal of Applied Econometrics*, 20, 445–465.

[47] McFadden, D.L., and P.A. Ruud (1994): "Estimation by Simulation," *Review of Economics and Statistics*, 76, 591–608.

[48] Nielsen, S. F. (2000a): "The Stochastic EM Algorithm: Estimation and Asymptotic Results," *Bernouilli*, 6(3): 457-489.

[49] Nielsen, S. F. (2000b): "On Simulated EM Algorithms," *Journal of Econometrics*, 96: 267-292.

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

[51] Schennach, S. (2014): "Entropic Latent Variable Integration via Simulation," *Econometrica*, 82(1), 345–385.

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

APPENDIX

A Identification

A.1 Discussion of Assumption 3

Part (i) in Assumption 3 requires that all densities under consideration be bounded. This imposes mild restrictions on the model's parameters. Part (ii) requires that $f_{Y_3|\eta,X}$ be non-identical at different values of η. This assumption will be satisfied if, for some τ in small open neighborhood $Q_{Y_3}(x,\eta_1,\tau) \neq Q_{Y_3}(x,\eta_2,\tau)$. In Example 2, part (i) requires strict monotonicity of quantile functions; that is: $x'\nabla\beta(\tau) + \eta\nabla\gamma(\tau) \geq \underline{c} > 0$, where $\nabla\xi(\tau)$ denotes the first derivative of $\xi(\cdot)$ evaluated at τ, while part (ii) holds if $\gamma(\tau) \neq 0$ for τ in some open neighborhood.

Part (iii) imposes a centered measure of location on $f_{Y_2|\eta,X=x}$. 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 Example 2 one convenient normalization is obtained by noting that:

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

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 and a rank condition is satisfied, $\overline{\beta}_1$ 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):

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

In a fully nonparametric setting and arbitrary t, to ensure that Assumption 3 (iii) holds for some period ($t = 1$, say) one can proceed as follows. First, let us define:

$$\widetilde{\eta}_i \equiv \mathbb{E}\left(Y_{i1}\,|\,\eta_i, X_{i1}\right).$$

Then, in every period t we have, provided $\eta \mapsto \mathbb{E}\left(Y_{i1}\,|\,\eta_i = \eta, X_{i1} = x_1\right)$ is invertible for almost all x_1:

$$Y_{it} = Q_Y\left(X_{it}, \eta_i, U_{it}\right) \equiv \widetilde{Q}_Y\left(X_{it}, X_{i1}, \widetilde{\eta}_i, U_{it}\right).$$

Estimating specifications of this form will deliver estimates of $\widetilde{Q}_Y$, from which the average marginal effects defined in Section 5 can be recovered as estimates of:

$$M_t(\tau) = \mathbb{E}\left[\frac{\partial Q_Y\left(X_{it}, \eta_i, \tau\right)}{\partial x_t}\right] = \mathbb{E}\left[\frac{\partial \widetilde{Q}_Y\left(X_{it}, X_{i1}, \widetilde{\eta}_i, \tau\right)}{\partial x_t}\right],$$

where $\partial\widetilde{Q}_Y/\partial x_t$ denotes the vector of partial derivatives of $\widetilde{Q}_Y$ with respect to its first $\dim(X_{it})$ arguments.

Part (iv) in Assumption 3 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 the only solution to $L_{Y_2|\eta,x}h = 0$ is $h = 0$. 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. Similarly as completeness, injectivity is a high-level condition; see for example Canay *et al.* (2012) for results on the testability of completeness assumptions.

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. Results by Hu and Shiu (2012) apply to the location-scale quantile model of Example 1. In this case, conditions that guarantee that $L_{Y_2|\eta,x}$ is injective involve the tail properties of the conditional density of Y_{i2} given η_i (and X_i) and its characteristic function.[16] Providing primitive conditions for injectivity/completeness in more general models, such as the linear quantile regression model of Example 2, is an interesting question but exceeds the scope of this paper.

A.2 Informal sketch of the arguments

We consider two setups in turn: the static model of Section 2, and the autoregressive first-order Markov model with exogenous regressors of Section 4. In both cases we provide a brief informal sketch of the identification argument.

Static model. Consider the static model of Subsection 2. For simplicity we leave the conditioning on covariates X_i implicit. Following Hu and Schennach (2008), we define several linear operators, which act on spaces of bounded functions. Let y_2 be one element in the support of Y_{i2}. 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 $h : y_1 \mapsto h(y_1)$, and provided integrals can be switched:

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

We thus have:

$$L_{Y_1,(y_2),Y_3} = L_{Y_3|\eta}\Delta_{(y_2)|\eta}L_{Y_1,\eta}, \quad y_2 - a.e. \tag{A1}$$

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

This yields a joint diagonalization system of operators, because, under suitable invertibility (i.e., injectivity) conditions, (A1) implies:

$$L_{Y_1,(y_2),Y_3}L^{-1}_{Y_1,(\widetilde{y}_2),Y_3} = L_{Y_3|\eta}\Delta_{(y_2)|\eta}\Delta^{-1}_{(\widetilde{y}_2)|\eta}L^{-1}_{Y_3|\eta}, \quad (y_2,\widetilde{y}_2)-a.e. \tag{A2}$$

The conditions of Hu and Schennach (2008)'s theorem then guarantee uniqueness of the solutions to (A2).

Dynamic autoregressive model. Let us now consider the dynamic autoregressive model of Section 4. As in Hu and Shum (2012) we define several operators. Let (y_2, y_3) be an element in the support of (Y_{i2}, Y_{i3}). 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:

$$L_{Y_1,(y_2),(y_3),Y_4} = L_{Y_4|(y_3),\eta}\Delta_{(y_3)|(y_2),\eta}L_{Y_1,(y_2),\eta}, \quad (y_2,y_3)-a.e. \tag{A3}$$

Hence, under suitable invertibility conditions:

$$L_{Y_1,(y_2),(y_3),Y_4}L^{-1}_{Y_1,(y_2),(\widetilde{y}_3),Y_4} = L_{Y_4|(y_3),\eta}\Delta_{(y_3)|(y_2),\eta}\Delta^{-1}_{(\widetilde{y}_3)|(y_2),\eta}L^{-1}_{Y_4|(\widetilde{y}_3),\eta}, \quad (y_2,y_3,\widetilde{y}_3)-a.e. \tag{A4}$$

Hu and Shum (2012), in particular in their Lemma 3, provide conditions for uniqueness of the solutions to (A4). Their conditions are closely related to the ones in Hu and Schennach (2008); see Assumption 3 in Subsection 2.3.

B Asymptotic results

B.1 Parametric inference

Here we rely on Nielsen's work to characterize the asymptotic distribution of $\widehat{\xi}^{(s)}$ in our model, where the optimization step is not likelihood-based but relies on different estimating equations. To do so, let us rewrite the moment restrictions in a compact notation:

$$\mathbb{E}[\Psi_i(\eta_i;\overline{\xi})] = 0,$$

where ξ (with true value $\overline{\xi}$) is a finite-dimensional parameter vector of the same dimension as Ψ. Equivalently, we have

$$\mathbb{E}\left[\int \Psi_i(\eta;\overline{\xi})f(\eta|W_i;\overline{\xi})d\eta\right] = 0,$$

where $W_i = (Y_i, X_i')'$.

The stochastic EM algorithm for this problem works as follows, based on an i.i.d. sample $(W_1, ..., W_N)$. Iteratively, one draws $\widehat{\xi}^{(s+1)}$ given $\widehat{\xi}^{(s)}$ in two steps:

1. For $i = 1, ..., N$, draw $\eta_i^{(1,s)}, ..., \eta_i^{(M,s)}$ from the posterior distribution $f(\eta_i | W_i; \widehat{\xi}^{(s)})$.[17]

2. Solve for $\widehat{\xi}^{(s+1)}$ in:

$$\sum_{i=1}^{N} \sum_{m=1}^{M} \Psi_i(\eta_i^{(m,s)}; \widehat{\xi}^{(s+1)}) = 0.$$

This results in a Markov Chain $(\widehat{\xi}^{(0)}, \widehat{\xi}^{(1)}, ...)$, which is ergodic under suitable conditions. Moreover, under conditions given in Nielsen (2000a), asymptotically as N tends to infinity the process $\sqrt{N}(\widehat{\xi}^{(s)} - \widehat{\xi})$ converges to a Gaussian autoregressive process conditional on almost every W-sequence, where $\widehat{\xi}$ solves the integrated moment restrictions:

$$\sum_{i=1}^{N} \int \Psi_i(\eta; \widehat{\xi}) f(\eta | W_i; \widehat{\xi}) d\eta = 0. \tag{B5}$$

In the rest of this section we characterize the unconditional asymptotic distribution of $\sqrt{N}(\widehat{\xi}^{(s)} - \overline{\xi})$. The derivations in this section are heuristic, and throughout we assume sufficient regularity conditions to justify all the steps.[18]

Using a conditional quantile representation we can write:

$$\eta_i^{(m,s)} = Q_{\eta|W}\left(W_i, V_i^{(m,s)}; \widehat{\xi}^{(s)}\right),$$

where $V_i^{(m,s)}$ are standard uniform draws, independent of each other and independent of W_i.

We thus have:

$$\sum_{i=1}^{N} \sum_{m=1}^{M} \Psi_i\left(Q_{\eta|W}\left(W_i, V_i^{(m,s)}; \widehat{\xi}^{(s)}\right); \widehat{\xi}^{(s+1)}\right) = 0.$$

Expanding around $\widehat{\xi}$, we obtain:

$$A\left(\widehat{\xi}^{(s+1)} - \widehat{\xi}\right) + B\left(\widehat{\xi}^{(s)} - \widehat{\xi}\right) + \varepsilon^{(s)} \quad = \quad o_p\left(N^{-\frac{1}{2}}\right), \tag{B6}$$

where:

$$A \equiv \left.\frac{\partial}{\partial \xi'}\right|_{\overline{\xi}} \mathbb{E}\left[\Psi_i\left(Q_{\eta|W}\left(W_i, V_i; \overline{\xi}\right); \xi\right)\right] = \left.\frac{\partial}{\partial \xi'}\right|_{\overline{\xi}} \mathbb{E}\left[\Psi_i\left(\eta_i; \xi\right)\right],$$

$$B \equiv \left.\frac{\partial}{\partial \xi'}\right|_{\overline{\xi}} \mathbb{E}\left[\Psi_i\left(Q_{\eta|W}\left(W_i, V_i; \xi\right); \overline{\xi}\right)\right] = \left.\frac{\partial}{\partial \xi'}\right|_{\overline{\xi}} \mathbb{E}\left[\int \Psi_i\left(\eta; \overline{\xi}\right) f\left(\eta | W_i; \xi\right) d\eta\right],$$

and:

$$\varepsilon^{(s)} \equiv \frac{1}{NM} \sum_{i=1}^{N} \sum_{m=1}^{M} \Psi_i\left(Q_{\eta|W}\left(W_i, V_i^{(m,s)}; \overline{\xi}\right); \overline{\xi}\right).$$

[17] For simplicity we consider the case where $\eta_i^{(1,s)}, ..., \eta_i^{(M,s)}$ are independent draws.

[18] Note that in our quantile model some of the moment restrictions involve derivatives of "check" functions, which are not smooth. This is however not central to the discussion that follows, as it does not affect the form of the asymptotic variance.

Note that:

$$A+B=\left.\frac{\partial}{\partial \xi'}\right|_{\overline{\xi}} \mathbb{E}\left[\int \Psi_i(\eta;\xi)f(\eta|W_i;\xi)d\eta\right].$$

The identification condition for the method-of-moments problem thus requires $A+B<0$, so $(-A)^{-1}B<I$. This implies that the autoregressive process $\sqrt{N}\left(\widehat{\xi}^{(s)}-\widehat{\xi}\right)$ is asymptotically stable. Conditionally on almost every W-sequence, $\sqrt{N}\left(\widehat{\xi}^{(s)}-\widehat{\xi}\right)$ is a stable Gaussian AR(1) process. We thus have:

$$\sqrt{N}\left(\widehat{\xi}^{(s)}-\widehat{\xi}\right)=\sum_{k=0}^{\infty}\left(-A^{-1}B\right)^k\left(-A^{-1}\right)\sqrt{N}\varepsilon^{(s-1-k)}+o_p(1). \tag{B7}$$

Moreover, $\sqrt{N}\varepsilon^{(s)}$ are asymptotically i.i.d. normal with zero mean and variance Σ/M, where:

$$\Sigma=\mathbb{E}\left[\Psi_i\left(\eta_i;\overline{\xi}\right)\Psi_i\left(\eta_i;\overline{\xi}\right)'\right].$$

Hence, conditionally on almost every W-sequence:

$$\sqrt{N}\left(\widehat{\xi}^{(s)}-\widehat{\xi}\right)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{V}_M\right),$$

where:

$$\mathcal{V}_M=\sum_{k=0}^{\infty}\left(-A^{-1}B\right)^k\left(-A^{-1}\right)\frac{\Sigma}{M}\left(-A^{-1}\right)'\left(\left(-A^{-1}B\right)^k\right)'.$$

Note that $\mathcal{V}_M$ can be recovered from the following matrix equation:

$$A^{-1}B\mathcal{V}_MB'(A^{-1})'=\mathcal{V}_M-A^{-1}\frac{\Sigma}{M}(A^{-1})',$$

which can be easily solved in vector form.

Finally, unconditionally we have by asymptotic independence:

$$\sqrt{N}\left(\widehat{\xi}^{(s)}-\overline{\xi}\right)=\sqrt{N}\left(\widehat{\xi}^{(s)}-\widehat{\xi}\right)+\sqrt{N}\left(\widehat{\xi}-\overline{\xi}\right)\xrightarrow{d}\mathcal{N}\left(0,\mathcal{V}+\mathcal{V}_M\right),$$

where $\mathcal{V}$ is the asymptotic variance of $\sqrt{N}\left(\widehat{\xi}-\overline{\xi}\right)$; that is:

$$\mathcal{V}=(A+B)^{-1}\Omega((A+B)^{-1})',$$

where $\Omega=\mathbb{E}\left[\left(\int\Psi_i(\eta;\overline{\xi})f(\eta|W_i;\overline{\xi})d\eta\right)\left(\int\Psi_i(\eta;\overline{\xi})f(\eta|W_i;\overline{\xi})d\eta\right)'\right]$.

B.2 Nonparametric consistency

Let $\overline{\xi}(\tau)=\left(\overline{\theta}(\tau)',\overline{\delta}(\tau)'\right)'$, and let $\varphi_i\left(\xi(\cdot),\tau\right)$ be the $(K_1+K_2)\times 1$ moment vector that corresponds to the integrated moment restrictions (20)-(21). Let $\|\cdot\|$ denote the Euclidean norm on $\mathbb{R}^{K_1+K_2}$, and let $\|\xi(\cdot)\|_\infty=\sup_{\tau\in(0,1)}\|\xi(\tau)\|$ denote the associated uniform norm.

Let $K=K_1+K_2$. The space $\mathcal{H}_K$ of functions $\xi(\cdot)$ contains differentiable functions whose first derivatives (component-wise) are bounded and Lipschitz on $(0,1)$. Moreover, there exists a $\underline{c}$ such that, for all $\tau_1<\tau_2$ and with probability one, $W_{it}(\eta_i)'\left(\theta(\tau_2)-\theta(\tau_1)\right)\geq\underline{c}(\tau_2-\tau_1)$ and $Z_i'(\delta(\tau_2)-\delta(\tau_1))\geq\underline{c}(\tau_2-\tau_1)$. This last requirement imposes strict monotonicity of the conditional quantile functions. These assumptions guarantee that the implied likelihood functions and posterior density of the individual effects are bounded from above and away from zero. Finally, all functions $\xi(\cdot)\in\mathcal{H}_K$ satisfy a location restriction as in Assumption 3 (*iii*).

To every function $\xi(\cdot) \in \mathcal{H}_K$ we associate an interpolating spline $\pi_L \xi(\cdot) \in \mathcal{H}_{KL}$. We use piecewise-linear splines on $(\tau_1, ..., \tau_L)$, as in Subsection 6.1. To simplify the analysis, we consider the case where quantile functions are constant on the tail intervals, so $\pi_L \xi(\tau) = \xi(\tau_1)$ for $\tau \in (0, \tau_1)$, and $\pi_L \xi(\tau) = \xi(\tau_L)$ for $\tau \in (\tau_L, 1)$. Moreover, the minimum and maximum of $L|\tau_{\ell+1} - \tau_\ell|$ are asymptotically bounded away from zero and infinity. As a result, $\|\xi(\cdot) - \pi_L \xi(\cdot)\|_\infty = O(\sqrt{K}/L)$, which we assume to tend to zero asymptotically.

Let us define:

$$Q_K(\xi(\cdot)) = \int_0^1 \|\mathbb{E}[\varphi_i(\xi(\cdot), \tau)]\|^2 d\tau,$$

and

$$\widehat{Q}_{KL}(\xi(\cdot)) = \frac{1}{L} \sum_{\ell=1}^{L} \left\| \frac{1}{N} \sum_{i=1}^{N} \varphi_i(\pi_L \xi(\cdot), \tau_\ell) \right\|^2.$$

We make the following high-level assumptions, which we will discuss below.

Assumption B1.

(i) *(identification) For all $\epsilon > 0$ there is a $c > 0$ such that, for all K_1, K_2, L:*

$$\inf_{\xi(\cdot) \in \mathcal{H}_K, \, \|\xi(\cdot) - \overline{\xi}(\cdot)\|_\infty > \epsilon} Q_K(\xi(\cdot)) \; > \; Q_K(\overline{\xi}(\cdot)) + c.$$

(ii) *(uniform convergence) As N, K_1, K_2, L tend to infinity:*

$$\sup_{\xi(\cdot) \in \mathcal{H}_K} \left| \widehat{Q}_{KL}(\xi(\cdot)) - Q_K(\xi(\cdot)) \right| = o_p(1).$$

Proposition B1. *(nonparametric consistency)*

Under Assumption B1, $\widehat{\xi}(\cdot)$ is uniformly consistent for $\overline{\xi}(\cdot)$ in the sense that (28) holds.

Proof. Let $\widetilde{\xi}(\cdot) \in \mathcal{H}_K$ such that $\widehat{\xi}(\cdot) = \pi_L \widetilde{\xi}(\cdot)$. We have $\left\|\widetilde{\xi}(\cdot) - \widehat{\xi}(\cdot)\right\|_\infty = o_p(1)$.

By definition of $\widehat{\xi}$ we have: $\widehat{Q}_{KL}\left(\widehat{\xi}(\cdot)\right) \leq \widehat{Q}_{KL}\left(\overline{\xi}(\cdot)\right)$. Let $\epsilon > 0$. By Assumption B1 (ii), and as $\widehat{Q}_{KL}\left(\widehat{\xi}(\cdot)\right) = \widehat{Q}_{KL}\left(\widetilde{\xi}(\cdot)\right)$:

$$Q_K\left(\widetilde{\xi}(\cdot)\right) \leq Q_K\left(\overline{\xi}(\cdot)\right) + o_p(1),$$

so, by Assumption B1 (i), $\left\|\widetilde{\xi}(\cdot) - \overline{\xi}(\cdot)\right\|_\infty \leq \epsilon$ with probability approaching one. This shows (28). ∎

Discussion of Assumption B1 (i). To provide intuition on the identification condition in Assumption B1 (i), consider the case where the posterior density $f(\eta|Y_i, X_i)$ is known. Consider the last K_2 elements of φ_i, the argument for the first K_1 elements being similar. Showing Assumption B1 (i) requires bounding the following quantity from below:

$$\Delta \equiv \int_0^1 \left\|\mathbb{E}\left[Z_i\left(\tau - F\left(Z_i'\delta(\tau)|Y_i, X_i\right)\right)\right]\right\|^2 - \left\|\mathbb{E}\left[Z_i\left(\tau - F\left(Z_i'\overline{\delta}(\tau)|Y_i, X_i\right)\right)\right]\right\|^2 d\tau.$$

Expanding yields:

$$\begin{aligned}\mathbb{E}\left[Z_i\left(\tau - F\left(Z_i'\delta(\tau)|Y_i, X_i\right)\right)\right] \;&=\; \mathbb{E}\left[Z_i\left(\tau - F\left(Z_i'\overline{\delta}(\tau)|Y_i, X_i\right)\right)\right] \\ &\quad - \mathbb{E}\left[Z_i Z_i' f\left(A_i(\tau; \delta)|Y_i, X_i\right)\right]\left(\delta(\tau) - \overline{\delta}(\tau)\right),\end{aligned}$$

where $A_i(\tau;\delta)$ lies between $Z_i'\delta(\tau)$ and $Z_i'\overline{\delta}(\tau)$. Now, $\mathbb{E}\left[Z_i\left(\tau - F\left(Z_i'\overline{\delta}(\tau)|Y_i, X_i\right)\right)\right] = o(1)$, provided the remainder R_η tends to zero sufficiently fast as K_2 increases. Moreover, if $f(\eta|Y_i, X_i)$ is bounded away from zero as well as from above, and if the eigenvalues of the Gram matrix $\mathbb{E}\left[Z_iZ_i'\right]$ are bounded away from zero as well as from above, then there exists a constant $\mu > 0$ such that:

$$\left\|\mathbb{E}\left[Z_iZ_i'f\left(A_i(\tau;\delta)|Y_i, X_i\right)\right]\left(\delta(\tau) - \overline{\delta}(\tau)\right)\right\|^2 \geq \mu\left\|\delta(\tau) - \overline{\delta}(\tau)\right\|^2.$$

Finally, suppose $\left\|\delta(\cdot) - \overline{\delta}(\cdot)\right\|_\infty > \epsilon$. Then by continuity of $\delta(\cdot) - \overline{\delta}(\cdot)$ there exists a non-empty interval (τ_1, τ_2) such that $\left\|\delta(\tau) - \overline{\delta}(\tau)\right\| > \epsilon$ for $\tau \in (\tau_1, \tau_2)$. Hence $\Delta > \mu\epsilon^2|\tau_2 - \tau_1| + o(1)$.

In the panel quantile models considered in this paper $f(\eta|Y_i, X_i; \xi(\cdot))$ depends on the unknown function $\xi(\cdot) = (\theta(\cdot)', \delta(\cdot)')'$. As we pointed out in Subsection 2.3, identification then depends on high-level conditions such as operator injectivity. Here we do not provide primitive conditions for Assumption B1 (i) to hold in this case.

Discussion of Assumption B1 (ii)**.** The uniform convergence condition in Assumption B1 (ii) will hold if the following conditions are satisfied:

$$\begin{aligned}
A &\equiv \sup_{\xi(\cdot)\in\mathcal{H}_K}\left|\frac{1}{L}\sum_{\ell=1}^{L}\left\|\frac{1}{N}\sum_{i=1}^{N}\varphi_i\left(\pi_L\xi(\cdot), \tau_\ell\right)\right\|^2 - \frac{1}{L}\sum_{\ell=1}^{L}\left\|\mathbb{E}\left[\varphi_i\left(\pi_L\xi(\cdot), \tau_\ell\right)\right]\right\|^2\right| = o_p(1),\\
B &\equiv \sup_{\xi(\cdot)\in\mathcal{H}_K}\left|\frac{1}{L}\sum_{\ell=1}^{L}\left\|\mathbb{E}\left[\varphi_i\left(\pi_L\xi(\cdot), \tau_\ell\right)\right]\right\|^2 - \frac{1}{L}\sum_{\ell=1}^{L}\left\|\mathbb{E}\left[\varphi_i\left(\xi(\cdot), \tau_\ell\right)\right]\right\|^2\right| = o(1),\\
C &\equiv \sup_{\xi(\cdot)\in\mathcal{H}_K}\left|\frac{1}{L}\sum_{\ell=1}^{L}\left\|\mathbb{E}\left[\varphi_i\left(\xi(\cdot), \tau_\ell\right)\right]\right\|^2 - \int_0^1\left\|\mathbb{E}\left[\varphi_i\left(\xi(\cdot), \tau\right)\right]\right\|^2 d\tau\right| = o(1).
\end{aligned}$$

The A quantity involves the difference between the empirical and population objective functions of the approximating parametric model. In the second term in B, the posterior density of individual effects depends on the entire function $\xi(\cdot)$, as opposed to its spline approximation $\pi_L\xi(\cdot)$. Lastly, the second term in C involves an integral on the unit interval, which needs to be compared to an average on the grid of τ_ℓ's.

A, B, C can be bounded by first establishing that φ_i is Lipschitz. Specifically, that there exist constants $C_1 > 0, C_2 > 0, \nu > 0$ such that, for all $\xi_1(\cdot), \xi_2(\cdot)$ in $\mathcal{H}_{KL}$ and τ_1, τ_2 in $(0,1)$:

$$\left\|\varphi_i\left(\xi_2(\cdot), \tau_2\right) - \varphi_i\left(\xi_1(\cdot), \tau_1\right)\right\| \leq C_1\sqrt{K}\left\|\xi_2(\cdot) - \xi_1(\cdot)\right\|_\infty^\nu + C_2\sqrt{K}|\tau_2 - \tau_1|. \tag{B8}$$

Consider the first K_1 elements of φ_i (the last K_2 elements having a similar structure):

$$\int\sum_{t=1}^{T}W_{it}(\eta)\psi_\tau(Y_{it} - W_{it}(\eta)'\theta(\tau))f(\eta|Y_i, X_i; \pi_L\xi(\cdot))d\eta.$$

To establish (B8), one may assume that $\eta \mapsto W_{it}(\eta)'\theta(\tau)$ is invertible almost surely[19] and that its inverse is Lipschitz in $\theta(\tau)$, and then use the expression of $f(\eta|Y_i, X_i; \pi_L\xi(\cdot))$, which involves the piecewise-linear expressions (41) and (42).

The $\pi_L\xi(\cdot)$ belong to a compact KL-dimensional space. Using (B8), it can be shown that $A = o_p(1)$ provided K/L^ν tends to zero and KL/N tends to zero. The latter condition arises as $\pi_L\xi(\cdot)$ is finite-dimensional, with dimension KL. Wei and Carroll (2009) establish this result formally for a related model, in a case where $K = O(1)$.

Next, extending (B8) to hold for $\xi_1(\cdot)$ and $\xi_2(\cdot)$ in $\mathcal{H}_K$, and using that $\|\xi(\cdot) - \pi_L\xi(\cdot)\|_\infty = o(1)$, yields $B = o(1)$ provided K/L tends to zero sufficiently fast. Lastly, again using (B8) but now for $\xi_1(\cdot) = \xi_2(\cdot)$, and using that $K/L^2 = o(1)$, yields $C = o(1)$.

[19] Such a condition requires that the conditional quantile function of outcomes be monotone in η_i.

C Extension: autocorrelated disturbances

To allow for autocorrelated errors in model (3)-(4) we replace Assumption 1 (*iii*) by:

Assumption C2. *(autocorrelated errors)*
$(U_{i1}, ..., U_{iT})$ *is distributed as a copula* $C(u_1, ..., u_T)$*, independent of* (X_i, η_i).

Nonparametric identification of the model (including the copula) can be shown under Markovian assumptions, as in the autoregressive model of Section 4. For estimation we let the copula depend on a finite-dimensional parameter ϕ, which we estimate along with all quantile parameters. The iterative estimation algorithm is then easily modified by adding an update in Step 2 (the "M"-step):

$$\widehat{\phi}^{(s+1)} = \underset{\phi}{\operatorname{argmax}} \sum_{i=1}^{N} \sum_{m=1}^{M} \ln \left[c \left(F \left(Y_{i1} | X_{i1}, \eta_i^{(m)}; \widehat{\xi}_A^{(s+1)} \right), ..., F \left(Y_{iT} | X_{iT}, \eta_i^{(m)}; \widehat{\xi}_A^{(s+1)} \right); \phi \right) \right], \tag{C9}$$

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 $\left(w_t(\eta)' \theta(\tau_1), w_t(\eta)' \theta(\tau_L)\right)$. Here F is a shorthand for $F_{Y_t|X_t,\eta}$.

The posterior density is then given by:

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

Lastly, note that the approach outlined here does not seem to easily generalize to allow for autocorrelated disturbances in autoregressive models (that is, for ARMA-type quantile regression models).

D Exponential modelling of the tails

For implementation, we use the following modelling for the splines in the extreme intervals indexed by $\lambda_1 > 0$ and $\lambda_L > 0$:

$$\begin{aligned} \theta(\tau) &= \theta(\tau_1) + \frac{\ln(\tau/\tau_1)}{\lambda_1} \iota_c, \quad \tau \leq \tau_1, \\ \theta(\tau) &= \theta(\tau_L) - \frac{\ln((1-\tau)/(1-\tau_L))}{\lambda_L} \iota_c, \quad \tau > \tau_L, \end{aligned}$$

where ι_c is a vector of zeros, with a one at the position of the constant term in $\theta(\tau)$. We adopt a similar specification for $\delta(\tau)$, with parameters $\lambda_1^\eta > 0$ and $\lambda_L^\eta > 0$. 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 $L = 1$, $\lambda_1 = 1 - \tau_1$, and $\lambda_L = \tau_L$.

The implied approximating period-t outcome density is then:

$$\begin{aligned} f_{Y_t|X_t,\eta}\left(y_t \mid x_t, \eta; \xi_A\right) \;=\; & \sum_{\ell=1}^{L-1} \frac{\tau_{\ell+1}-\tau_\ell}{w_t\left(\eta\right)'\left[\theta\left(\tau_{\ell+1}\right)-\theta\left(\tau_\ell\right)\right]} \mathbf{1}\left\{w_t\left(\eta\right)'\theta\left(\tau_\ell\right) < y_t \leq w_t\left(\eta\right)'\theta\left(\tau_{\ell+1}\right)\right\} \\ & +\tau_1\lambda_1 e^{\lambda_1\left(y_t - w_t(\eta)'\theta(\tau_1)\right)}\mathbf{1}\left\{y_t \leq w_t\left(\eta\right)'\theta\left(\tau_1\right)\right\} \\ & +\left(1-\tau_L\right)\lambda_L e^{-\lambda_L\left(y_t - w_t(\eta)'\theta(\tau_L)\right)}\mathbf{1}\left\{y_t > w_t\left(\eta\right)'\theta\left(\tau_L\right)\right\}. \end{aligned}$$

Similarly, the approximating density of individual effects is:

$$\begin{aligned} f_{\eta|X}\left(\eta \mid x; \xi_B\right) \;=\; & \sum_{\ell=1}^{L-1} \frac{\tau_{\ell+1}-\tau_\ell}{z'\left[\delta\left(\tau_{\ell+1}\right)-\delta\left(\tau_\ell\right)\right]} \mathbf{1}\left\{z'\delta\left(\tau_\ell\right) < \eta \leq z'\delta\left(\tau_{\ell+1}\right)\right\} \\ & +\tau_1\lambda_1^\eta e^{\lambda_1^\eta\left(\eta - z'\delta(\tau_1)\right)}\mathbf{1}\left\{\eta \leq z'\delta\left(\tau_1\right)\right\} \\ & +\left(1-\tau_L\right)\lambda_L^\eta e^{-\lambda_L^\eta\left(\eta - z'\delta(\tau_L)\right)}\mathbf{1}\left\{\eta > z'\delta\left(\tau_L\right)\right\}. \end{aligned}$$

Update rules for exponential parameters. We adopt a likelihood approach to update the parameters $\lambda_1, \lambda_L, \lambda_1^\eta, \lambda_L^\eta$. This yields the following moment restrictions:

$$\overline{\lambda}_1^\eta = \frac{-\mathbb{E}\left[\int \mathbf{1}\left\{\eta \leq Z_i'\overline{\delta}(\tau_1)\right\} f(\eta|Y_i, X_i; \overline{\xi})d\eta\right]}{\mathbb{E}\left[\int \left(\eta - Z_i'\overline{\delta}(\tau_1)\right)\mathbf{1}\left\{\eta \leq Z_i'\overline{\delta}(\tau_1)\right\} f(\eta|Y_i, X_i; \overline{\xi})d\eta\right]},$$

and:

$$\overline{\lambda}_L^\eta = \frac{\mathbb{E}\left[\int \mathbf{1}\left\{\eta > Z_i'\overline{\delta}(\tau_L)\right\} f(\eta|Y_i, X_i; \overline{\xi})d\eta\right]}{\mathbb{E}\left[\int \left(\eta - Z_i'\overline{\delta}(\tau_L)\right)\mathbf{1}\left\{\eta > Z_i'\overline{\delta}(\tau_L)\right\} f(\eta|Y_i, X_i; \overline{\xi})d\eta\right]},$$

with similar equations for λ_1, λ_L.

Hence the update rules in Step 2 of the algorithm (the "M"-step):

$$\widehat{\lambda}_1^{\eta,(s+1)} = \frac{-\sum_{i=1}^N\sum_{m=1}^M \mathbf{1}\left\{\eta_i^{(m)} \leq Z_i'\widehat{\delta}(\tau_1)^{(s)}\right\}}{\sum_{i=1}^N\sum_{m=1}^M \left(\eta_i^{(m)} - Z_i'\widehat{\delta}(\tau_1)^{(s)}\right)\mathbf{1}\left\{\eta_i^{(m)} \leq Z_i'\widehat{\delta}(\tau_1)^{(s)}\right\}},$$

and:

$$\widehat{\lambda}_L^{\eta,(s+1)} = \frac{\sum_{i=1}^N\sum_{m=1}^M \mathbf{1}\left\{\eta_i^{(m)} > Z_i'\widehat{\delta}(\tau_L)^{(s)}\right\}}{\sum_{i=1}^N\sum_{m=1}^M \left(\eta_i^{(m)} - Z_i'\widehat{\delta}(\tau_L)^{(s)}\right)\mathbf{1}\left\{\eta_i^{(m)} > Z_i'\widehat{\delta}(\tau_L)^{(s)}\right\}}.$$

E Monte Carlo illustration

The data generating process is as follows:

$$Y_{it} = \beta_0\left(U_{it}\right) + \beta_1\left(U_{it}\right)X_{1it} + \beta_2\left(U_{it}\right)X_{2it} + \beta_3\left(U_{it}\right)X_{3it} + \gamma\left(U_{it}\right)\eta_i,$$

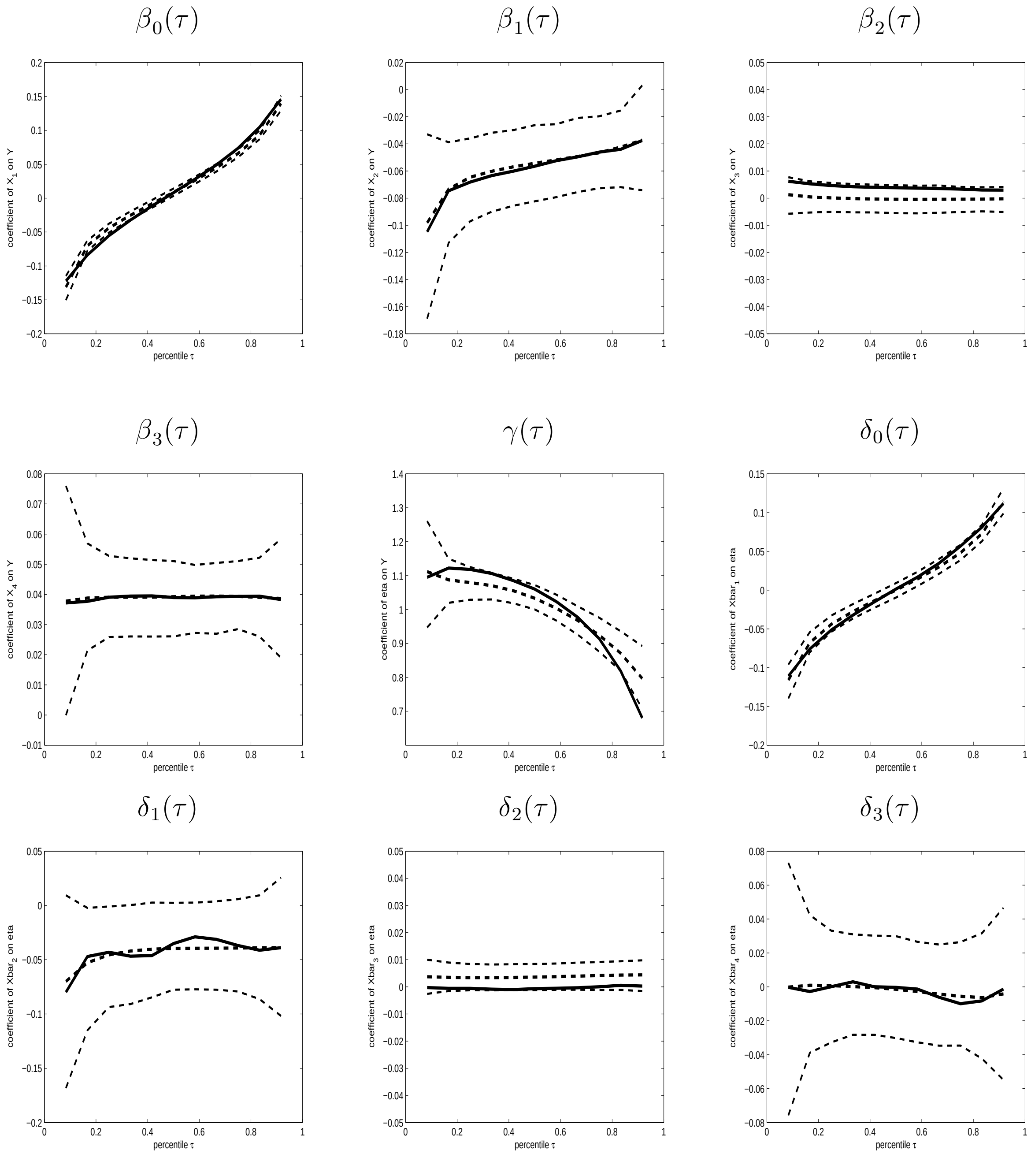
and:

$$\eta_i = \delta_0\left(V_i\right) + \delta_1\left(V_i\right)\overline{X}_{1i} + \delta_2\left(V_i\right)\overline{X}_{2i} + \delta_3\left(V_i\right)\overline{X}_{3i}.$$

The covariates X_{i1} (smoking status), X_{i2} (age), and X_{i3} (gender) are taken from the data set of the empirical illustration. $T = 3$, and we extract a random subsample of 1000 mothers from the original data set. The true parameter values correspond to estimates on the full sample. Parameters β's, γ's and δ's are taken to be piecewise-linear on an equidistant grid with $L = 11$ knots, with exponential specifications in the tails of intercept coefficients. For computation we use the same method as in the application to select starting values, and we let the EM algorithm run for 100 iterations, with 100 random-walk Metropolis-Hastings draws within each iteration, reporting averages over the last 50 iterations. We report the results of 500 simulations.

The results in Figure E1 show moderate biases and relatively precise estimates. For example, the confidence intervals of the quantile parameters $\beta_1(\tau)$ corresponding to the effect of smoking are quite tight, even though the sample size is about 12 times smaller than the one of the application. Overall the results provide encouraging evidence on the finite sample performance of the estimator.

Figure E1: Monte Carlo results



Note: Data generating process with $L = 11$ knots, $N = 1000$, $T = 3$. The x-axis shows τ percentiles. True parameter values (solid lines), Monte Carlo means (thick dashed lines), and 95% pointwise confidence intervals (thin dashed lines). 500 simulations.