Nonlinear Panel Data Estimation
via Quantile Regressions*

Manuel Arellano   Stéphane Bonhomme
CEMFI, Madrid   University of Chicago

First draft: February 2013
This draft: July 2015

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 children's birthweights completes the paper.

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

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 thorough account of these methods. In this work we show that quantile regression can also be successfully applied to panel data.

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 to build flexible models for the dependence of unobserved heterogeneity on exogenous covariates or initial conditions, and for the feedback processes of covariates in models with general predetermined regressors.

We consider classes of panel data models with continuous outcomes that satisfy conditional independence restrictions, but are otherwise nonparametric. 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 general nonlinear panel models. For this purpose, we specify outcomes Y_{it} as a function of covariates X_{it} and latent heterogeneity η_i as:

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

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

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

where $U_{i1}, ..., U_{iT}, V_i$ are independent uniform random variables, and g's and h's belong to some family 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 for flexible patterns of 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. Moreover, while (2) is 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 $g_k(X_{it}, \eta_i)$. 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 solving a quantile regression averaged over imputed values.

For the imputed values to be valid they have to be draws from the 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 a grid of 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.

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 method-of-moments estimators 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, make each parameter update step 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 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 (2010) 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 (flexible) 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

---

[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.

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 seminal work by 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 and dynamic models, and discuss identification. In Section 3 we present our estimation method and study some of its properties. In Section 4 we present the empirical illustration. Lastly, we conclude in Section 5. Proofs and further discussion are contained in the Appendix.

2   Nonlinear quantile models for panel data

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

2.1   Static models

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:

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

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 the next subsection. 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 $\tau \in (0, 1)$, $Q_Y(X_{it}, \eta_i, \tau)$ 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. Part (iii) defines the static version of the model, where U_{it} are assumed to be independent over time. In the next subsection 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:

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

---

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

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

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

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

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

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

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

---

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

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

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

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

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

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

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):

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

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:

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

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:

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

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.[4]

Series and smooth coefficients approaches. In this paper, our approach to estimation will be based on series specifications of the form (1) and (2), which will generalize Examples 1-2-3. As the number of series terms increases these specifications provide flexible approximations to the conditional quantile functions in (3) and (4).

A different approach would be to allow for smooth coefficients based on local polynomial approximations, for example:[5]

$$Y_{it} = X'_{it}\beta\left(U_{it}, \eta_i\right) + \gamma\left(U_{it}, \eta_i\right),\tag{12}$$

where, for any given point η^*:

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

with $\beta\left(\tau, \eta^*\right) = b_{\tau 0}\left(\eta^*\right)$ and $\gamma\left(\tau, \eta^*\right) = c_{\tau 0}\left(\eta^*\right)$. Although we will not analyze this setup in detail, one could adapt our estimation approach by using locally weighted check function (Chaudhuri, 1991, Cai and Xu, 2008).

2.2 Dynamic models

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:

$$Y_{it} = Q_Y\left(Y_{i,t-1}, X_{it}, \eta_i, U_{it}\right), \quad i = 1, ..., N, \quad t = 2, ..., T.\tag{13}$$

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 (13) 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

---

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

[5]See Belloni, Chernozhukov and Fernández-Val (2012) and Qu and Yoon (2011) for conditional quantile estimation based on series expansions and nonparametric kernels, respectively.

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 (13), 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:

$$\eta_i = Q_\eta (Y_{i1}, X_i, V_i), \quad i = 1, ..., N, \tag{14}$$

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:

$$X_{it} = Q_X (Y_{i,t-1}, X_{i,t-1}, \eta_i, A_{it}), \quad i = 1, ..., N, \quad t = 2, ..., T. \tag{15}$$

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:

$$\eta_i = Q_\eta (Y_{i1}, X_{i1}, V_i), \quad i = 1, ..., N. \tag{16}$$

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

Assumption 3. *(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 (15) 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 $X_{it} = (X_{1it}, X_{2it})'$:

$$
\begin{aligned}
X_{1it} &= Q_{X_1}\left(Y_{i,t-1}, X_{1i,t-1}, X_{2i,t-1}, \eta_i, A_{1it}\right), \\
X_{2it} &= Q_{X_2}\left(Y_{i,t-1}, X_{1it}, X_{1i,t-1}, X_{2i,t-1}, \eta_i, A_{2it}\right),
\end{aligned}
\tag{17}
$$

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:

$$
Y_{it} = \rho\left(U_{it}\right) Y_{i,t-1} + X'_{it}\beta\left(U_{it}\right) + \eta_i \gamma\left(U_{it}\right).
\tag{18}
$$

Model (18) differs from the more general model studied in Galvao (2011):

$$
Y_{it} = \rho\left(U_{it}\right) Y_{i,t-1} + X'_{it}\beta\left(U_{it}\right) + \eta_i\left(U_{it}\right).
\tag{19}
$$

Similarly as in (7), and in contrast with the models introduced in this paper, the presence of the functional heterogeneity term $\eta_i\left(\tau\right)$ makes fixed-T consistent estimation problematic in (19).

An extension of (18) is:

$$
Y_{it} = h\left(Y_{i,t-1}\right)' \rho\left(U_{it}\right) + X'_{it}\beta\left(U_{it}\right) + \eta_i \gamma\left(U_{it}\right), \quad t = 2, ..., T,
\tag{20}
$$

where h is a univariate function. For example, when $h(y) = |y|$ model (20) 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 (18) 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 (17).

2.3 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 heterogeneous, equal to $\beta(\tau) + \gamma_2(\tau)\eta_{i2}$, and $M(\tau) = \beta(\tau) + \gamma_2(\tau)\mathbb{E}[\eta_{i2}]$. 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, $\mathbb{E}[\partial Q_Y(Y_{i,t-1}, X_{it}, \eta_i, \tau)/\partial y]$, 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 4, 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 off 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 will be identified for this model. 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:[6]

$$F_{Y_{it}(d)}(y) = \mathbb{E}\left[\int_0^1 \mathbf{1}\left\{Q_Y(d, X_{it}, \eta_i, \tau) \leq y\right\} d\tau\right], \quad d \in \{0, 1\}. \tag{21}$$

We next discuss conditions under which all these quantities are nonparametrically identified for a fixed number of time periods.

2.4   Nonparametric identification

The panel data models introduced above satisfy conditional independence restrictions. In the class of static models of Subsection 2.1, period-specific outcomes $Y_{i1}, ..., Y_{iT}$ are mutually independent conditional on exogenous covariates and individual heterogeneity X_i, η_i. The dynamic models of Subsection 2.2 satisfy Markovian independence restrictions.

---

[6]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.

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 results in Hu and Schennach (2008) and Hu and Shum (2012) can be used to show nonparametric identification of the nonlinear panel data models that we consider in this paper.

Static setup. 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:

$$f_{Y_1, Y_2, Y_3 | \eta, X} (y_1, y_2, y_3 \mid \eta, x) = f_{Y_1 | \eta, X} (y_1 \mid \eta, x) \, f_{Y_2 | \eta, X} (y_2 \mid \eta, x) \, f_{Y_3 | \eta, X} (y_3 \mid \eta, x) . \qquad (22)$$

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

$$f_{Y_1, Y_2, Y_3 | X} (y_1, y_2, y_3 \mid x) =$$
$$\int f_{Y_1 | \eta, X} (y_1 \mid \eta, x) \, f_{Y_2 | \eta, X} (y_2 \mid \eta, x) \, f_{Y_3 | \eta, X} (y_3 \mid \eta, x) \, f_{\eta | X} (\eta \mid x) \, d\eta. \qquad (23)$$

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

Assumption 4. *(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 $\eta_1 \neq \eta_2$: $\Pr \left[f_{Y_3 | \eta, X}(Y_{i3} | \eta_1, x) \neq f_{Y_3 | \eta, X}(Y_{i3} | \eta_2, x) \mid X_i = x \right] > 0.$

(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 4 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:

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

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 4 (iii):

$$\overline{\beta}_0 = \int_0^1 \beta_0\left(\tau\right)d\tau = 0, \quad \text{and} \quad \overline{\gamma} = \int_0^1 \gamma\left(\tau\right)d\tau = 1. \tag{24}$$

We will use (24) in our empirical implementation.[7]

In a fully nonparametric setting and arbitrary t, to ensure that Assumption 4 (iii) holds for some period ($t = 1$, say) one can proceed as follows. First, let us define:

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

---

[7]In fact, Assumption 4 (iii) is also implied by $(\overline{\beta}_0, \overline{\gamma}_0) = (0, 1)$ in the following model with first-order interactions, a version of which we estimate in the empirical application:

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

Then, in every period t we have, provided $\eta \mapsto \mathbb{E}\left(Y_{i1} \mid \eta_i = \eta, X_{i1} = x_1\right)$ 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 average marginal effects 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 4 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.[8] 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.

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 D.

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

Let Assumptions 1, 2, and 4 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.

---

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

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

Models with multiple effects.   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.

Markovian dynamics.   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:

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

(25)

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 D.[9]

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

---

[9] In the dynamic model (20), it follows from Hu and Shum (2012)'s analysis that one can rely on (24) as in the static case, provided the averages across τ values of the coefficients of exogenous regressors and lagged outcome are identified based on:

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

3 Quantile regression estimators

In this section we introduce our estimation strategy and discuss several of its statistical properties. We first analyze a class of static quantile models in detail, and then show how the methodology can be extended to dynamic settings.

3.1 Model specification and moment restrictions

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

$$Q_Y(X_{it}, \eta_i, \tau) = W_{it}(\eta_i)' \theta(\tau). \tag{26}$$

In (26) the vector $W_{it}(\eta_i)$ 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 one takes $W_{it}(\eta_i) = (X'_{it}, \eta_i)'$. A more flexible approach is to use a series specification of the quantile function as in (1), and to set $W_{it}(\eta_i) = (g_1(X_{it}, \eta_i), ..., g_{K_1}(X_{it}, \eta_i))'$ for a set of K_1 functions $g_1, ..., g_{K_1}$. In practice one may use orthogonal polynomials, wavelets or splines, among other choices; see Chen (2007) for a comprehensive survey of sieve methods.

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

$$Q_\eta(X_i, \tau) = Z'_i \delta(\tau). \tag{27}$$

In (27) 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:

$$f_{\eta|Y,X}(\eta \mid y, x; \theta(\cdot), \delta(\cdot)) = \frac{\prod_{t=1}^{T} f_{Y_t|X_t, \eta}(y_t \mid x_t, \eta; \theta(\cdot)) f_{\eta|X}(\eta \mid x; \delta(\cdot))}{\int \prod_{t=1}^{T} f_{Y_t|X_t, \eta}(y_t \mid x_t, \widetilde{\eta}; \theta(\cdot)) f_{\eta|X}(\widetilde{\eta} \mid x; \delta(\cdot)) d\widetilde{\eta}}, \tag{28}$$

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):

$$\rho_\tau(u) = (\tau - \mathbf{1}\{u < 0\})u, \qquad \psi_\tau(u) = \nabla \rho(u).$$

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:

$$\mathbb{E}\left[\sum_{t=1}^{T} W_{it}\left(\eta_i\right)\psi_\tau\left(Y_{it} - W_{it}\left(\eta_i\right)'\theta\left(\tau\right)\right)\right] = 0, \tag{29}$$

and:

$$\mathbb{E}\left[Z_i\psi_\tau\left(\eta_i - Z_i'\delta\left(\tau\right)\right)\right] = 0. \tag{30}$$

Indeed, (29) is the first-order condition associated with the infeasible population quantile regression of Y_{it} on $W_{it}\left(\eta_i\right)$. Similarly, (30) corresponds to the infeasible quantile regression of η_i on Z_i.

Applying the law of iterated expectations to (29) and (30), respectively, we obtain the following integrated moment restrictions, for all $\tau \in (0,1)$:

$$\mathbb{E}\left[\int\left(\sum_{t=1}^{T} W_{it}\left(\eta\right)\psi_\tau\left(Y_{it} - W_{it}\left(\eta\right)'\theta\left(\tau\right)\right)\right)f\left(\eta \mid Y_i, X_i; \theta\left(\cdot\right), \delta\left(\cdot\right)\right)d\eta\right] = 0, \tag{31}$$

and:

$$\mathbb{E}\left[\int\left(Z_i\psi_\tau\left(\eta - Z_i'\delta\left(\tau\right)\right)\right)f\left(\eta \mid Y_i, X_i; \theta\left(\cdot\right), \delta\left(\cdot\right)\right)d\eta\right] = 0, \tag{32}$$

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 (31)-(32) that, if the posterior density of the individual effects were known, then estimating the model's parameters could be done using two simple linear quantile regressions, weighted by the posterior density. However, as the notation makes clear, the posterior density in (28) depends on the entire processes $\theta\left(\cdot\right)$ and $\delta\left(\cdot\right)$. Specifically we have, for absolutely continuous conditional densities of outcomes and individual effects:

$$f_{Y_t|X_t,\eta}\left(y_t \mid x_t, \eta; \theta\left(\cdot\right)\right) = \lim_{\epsilon \to 0}\frac{\epsilon}{w_t\left(\eta\right)'\left[\theta\left(u_t + \epsilon\right) - \theta\left(u_t\right)\right]}, \tag{33}$$

and:

$$f_{\eta|X}\left(\eta \mid x; \delta\left(\cdot\right)\right) = \lim_{\epsilon \to 0}\frac{\epsilon}{z'\left[\delta\left(v + \epsilon\right) - \delta\left(v\right)\right]}, \tag{34}$$

where u_t and v are defined by:

$$w_t\left(\eta\right)'\theta\left(u_t\right) = y_t, \quad \text{and:} \quad z'\delta\left(v\right) = \eta.$$

18

Equations (33) and (34) 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 $\theta(\tau)$ and $\delta(\tau)$ in (31)-(32) 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 $\theta(\cdot)$ and $\delta(\cdot)$ using splines, with L knots $0 < \tau_1 < \tau_2 < ... < \tau_L < 1$. A practical possibility is to use piecewise-linear splines as in Wei and Carroll, but other choices are possible, such as cubic splines or shape-preserving B-splines. When using interpolating splines, the approximation argument requires suitable smoothness assumptions on $\theta(\tau)$ and $\delta(\tau)$ as functions of $\tau \in (0,1)$. For fixed L, the spline specification may be seen as an approximation to the underlying quantile functions.

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

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

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

$$f(\eta \mid y, x; \xi) = \frac{\prod_{t=1}^{T} f_{Y_t|X_t,\eta}(y_t \mid x_t, \eta; \xi_A) f_{\eta|X}(\eta \mid x; \xi_B)}{\int \prod_{t=1}^{T} f_{Y_t|X_t,\eta}(y_t \mid x_t, \widetilde{\eta}; \xi_A) f_{\eta|X}(\widetilde{\eta} \mid x; \xi_B) \, d\widetilde{\eta}}. \tag{35}$$

The approximating densities take particularly simple forms 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 4.1 below.

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

$$\mathbb{E}\left[\int\left(\sum_{t=1}^{T} W_{it}(\eta)\,\psi_{\tau_\ell}\left(Y_{it} - W_{it}(\eta)'\,\theta(\tau_\ell)\right)\right)f(\eta \mid Y_i, X_i; \xi)\,d\eta\right] = 0, \qquad (36)$$

and:

$$\mathbb{E}\left[\int\left(Z_i\psi_{\tau_\ell}(\eta - Z_i'\delta(\tau_\ell))\right)f(\eta \mid Y_i, X_i; \xi)\,d\eta\right] = 0. \qquad (37)$$

3.2 Estimation algorithm

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

$$\widehat{\theta}(\tau_\ell) = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N}\int\left(\sum_{t=1}^{T}\rho_{\tau_\ell}\left(Y_{it} - W_{it}(\eta)'\,\theta\right)\right)f\left(\eta \mid Y_i, X_i; \widehat{\xi}\right)d\eta, \qquad (38)$$

$$\widehat{\delta}(\tau_\ell) = \underset{\delta}{\operatorname{argmin}} \sum_{i=1}^{N}\int\rho_{\tau_\ell}(\eta - Z_i'\delta)\,f\left(\eta \mid Y_i, X_i; \widehat{\xi}\right)d\eta, \qquad (39)$$

where $\rho_\tau(\cdot)$ is the check function, and where $f(\eta \mid Y_i, X_i; \xi)$ is given by (35). Note that the first-order conditions of (38)-(39) are the sample analogs of the moment restrictions (36)-(37) of the approximating model.

To solve the fixed-point problem (38)-(39) we proceed in an iterative fashion. Starting with initial parameter values $\widehat{\xi}^{(0)}$, one possibility is to iterate the following two steps until numerical convergence:

1. Compute the posterior density:

$$\widehat{f}_i^{(s)}(\eta) = f\left(\eta \mid Y_i, X_i; \widehat{\xi}^{(s)}\right). \qquad (40)$$

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

$$\widehat{\theta}(\tau_\ell)^{(s+1)} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N}\int\left(\sum_{t=1}^{T}\rho_{\tau_\ell}\left(Y_{it} - W_{it}(\eta)'\,\theta\right)\right)\widehat{f}_i^{(s)}(\eta)d\eta, \qquad (41)$$

$$\widehat{\delta}(\tau_\ell)^{(s+1)} = \underset{\delta}{\operatorname{argmin}} \sum_{i=1}^{N}\int\rho_{\tau_\ell}(\eta - Z_i'\delta)\,\widehat{f}_i^{(s)}(\eta)d\eta. \qquad (42)$$

This sequential method-of-moment method is related to, but different from, the standard EM algorithm (Dempster *et al.*, 1977). Similarly 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 "M"-step of our algorithm (that is, in equations (41)-(42)) estimation is not based on a likelihood function, but on the check function of quantile regression.

Proceeding in this way has two major computational advantages compared to maximizing the full likelihood of the approximating model. Firstly, as opposed to the likelihood function, which is a complicated function of all quantile regression coefficients, the problem in (41)-(42) nicely decomposes into L different τ_ℓ-specific subproblems. Secondly, using the check function yields a globally convex objective function in each "M"-step.

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.[10] Secondly, whereas conditions for numerical convergence of ordinary EM are available in the literature (e.g., Wu, 1983), formal proofs of convergence of sequential algorithms such as ours seem difficult to establish.[11]

Simulation-based algorithm. In practice we use a simulation-based approach in the first step of the estimation algorithm. This allows us to replace the integrals in (41)-(42) 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 $i = 1, ..., N$, draw M values $\eta_i^{(1)},...,\eta_i^{(M)}$ from the posterior density $\widehat{f}_i^{(s)}$ given by (40).*

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

$$\widehat{\theta}\left(\tau_\ell\right)^{(s+1)} = \operatorname*{argmin}_{\theta} \sum_{i=1}^{N} \sum_{m=1}^{M} \sum_{t=1}^{T} \rho_{\tau_\ell}\left(Y_{it} - W_{it}\left(\eta_i^{(m)}\right)' \theta\right),$$

$$\widehat{\delta}\left(\tau_\ell\right)^{(s+1)} = \operatorname*{argmin}_{\delta} \sum_{i=1}^{N} \sum_{m=1}^{M} \rho_{\tau_\ell}\left(\eta_i^{(m)} - Z_i'\delta\right).$$

---

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

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

In step one (the "E" step) in this algorithm 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). 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 Fernandez-Val (2010). In Step 2 (the "M" step) we run $2L$ ordinary quantile regressions, where the simulated values of the individual effects are treated, in turn, as covariates and dependent variables.

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). Below, building on Nielsen's work we analyze the statistical properties of estimators based on fixed-M and large-M versions of the algorithm.

3.3 Estimation in dynamic models

The estimation algorithm of the previous section can be directly modified to deal with autoregressive models with strictly exogenous covariates. Consider a linear specification of the quantile functions (13) and (14), 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:

$$f\left(\eta \mid y, x; \xi\right) = \frac{\prod_{t=2}^{T} f_{Y_t \mid Y_{t-1}, X_t, \eta}\left(y_t \mid y_{t-1}, x_t, \eta; \xi_A\right) f_{\eta \mid Y_1, X}\left(\eta \mid y_1, x; \xi_B\right)}{\int \prod_{t=1}^{T} f_{Y_t \mid Y_{t-1}, X_t, \eta}\left(y_t \mid y_{t-1}, x_t, \widetilde{\eta}; \xi_A\right) f_{\eta \mid Y_1, X}\left(\widetilde{\eta} \mid y_1, x; \xi_B\right) d\widetilde{\eta}}. \tag{43}$$

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:

$$f(\eta \mid y, x; \xi) = \frac{f_{W_2,...,W_T}(w_2, ..., w_T \mid w_1, \eta) f_{\eta \mid W_1}(\eta \mid w_1)}{\int f_{W_2,...,W_T}(w_2, ..., w_T \mid w_1, \eta) f_{\eta \mid W_1}(\eta \mid w_1) d\eta}$$

$$= \frac{f_{\eta \mid W_1}(\eta \mid w_1; \xi_B) \prod_{t=2}^{T} f_{Y_t \mid Y_{t-1}, X_t, \eta}(y_t \mid y_{t-1}, x_t, \eta; \xi_A) f_{X_t \mid W^{t-1}, \eta}(x_t \mid w^{t-1}, \eta; \xi_C)}{\int f_{\eta \mid W_1}(\widetilde{\eta} \mid w_1; \xi_B) \prod_{t=2}^{T} f_{Y_t \mid Y_{t-1}, X_t, \eta}(y_t \mid y_{t-1}, x_t, \widetilde{\eta}; \xi_A) f_{X_t \mid W^{t-1}, \eta}(x_t \mid w^{t-1}, \widetilde{\eta}; \xi_C) d\widetilde{\eta}},$$

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:

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

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

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

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 (15) 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:

$$X_{1it} = W_{1it}(\eta_i) \mu_1(A_{1it}),$$

$$\cdots \quad \cdots \quad \cdots$$

$$X_{Pit} = W_{Pit}(\eta_i) \mu_P(A_{Pit}), \tag{44}$$

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 (13) specified as a linear quantile regression, the model of the feedback process (44), and the model of individual effects (16), 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.

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

Assumption 5. *(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. 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):

$$\widehat{\phi}^{(s+1)} = \underset{\phi}{\operatorname{argmax}} \sum_{i=1}^{N} \sum_{m=1}^{M} \ln \left[c \left(F \left(Y_{i1} | X_{i1}, \eta_i^{(m)}; \widehat{\xi}_A^{(s+1)} \right), ..., F \left(Y_{iT} | X_{iT}, \eta_i^{(m)}; \widehat{\xi}_A^{(s+1)} \right); \phi \right) \right], \tag{45}$$

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 $w_t(\eta)' \theta(\tau_\ell) < y_t \le w_t(\eta)' \theta(\tau_{\ell+1})$:

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

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).

3.4 Asymptotic properties

We now discuss the asymptotic properties of the estimation algorithm. Throughout, T is fixed while N tends to infinity. Although we focus on a static model for concreteness, similar arguments apply to dynamic models with exogenous or predetermined covariates.

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:

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

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:

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

$\mathcal{V}$ being the asymptotic variance of the method-of-moments estimator based on the integrated moment restrictions (36)-(37).

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 $\widehat{\xi} = (\widetilde{\theta}', \widetilde{\delta}')'$ based on the static model and the integrated moment restrictions (36)-(37). 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:

$$\sup_{\tau \in (0,1)} \left\| \widehat{\xi}(\tau) - \overline{\xi}(\tau) \right\| = o_p(1), \tag{46}$$

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. 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.[12] 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. We leave a detailed study of the asymptotic properties of our estimator as N, K_1, K_2, and L tend to infinity to future work.

4   Empirical application

In this section we present an empirical illustration to the link between mother inputs such as smoking and children's birthweights. We start by discussing how we implement the estimation algorithm in practice. In Appendix C we report the results of a small Monte Carlo illustration.

---

[12]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).

4.1 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 $\ell = 1, ..., L - 1$:

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

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

$$
f_{Y_t|X_t,\eta}\left(y_t \mid x_t, \eta; \xi_A\right) = \frac{\tau_{\ell+1} - \tau_\ell}{w_t\left(\eta\right)'\left[\theta\left(\tau_{\ell+1}\right) - \theta\left(\tau_\ell\right)\right]} \quad \text{if } w_t\left(\eta\right)'\theta\left(\tau_\ell\right) < y_t \le w_t\left(\eta\right)'\theta\left(\tau_{\ell+1}\right), \tag{47}
$$

$$
f_{\eta|X}\left(\eta \mid x; \xi_B\right) = \frac{\tau_{\ell+1} - \tau_\ell}{z'\left[\delta\left(\tau_{\ell+1}\right) - \delta\left(\tau_\ell\right)\right]} \quad \text{if } z'\delta\left(\tau_\ell\right) < \eta \le z'\delta\left(\tau_{\ell+1}\right), \tag{48}
$$

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

Tail intervals. In order to model quantile functions in the intervals $(0, \tau_1)$ and $(\tau_L, 1)$ one may assume, following Wei and Carroll (2009), that $\theta(\cdot)$ and $\delta(\cdot)$ are constant on these intervals, so the implied distribution functions have mass points at the two ends of the support. In Appendix A we outline a different, 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.

4.2 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.

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.

Figure 1: Quantile effects of smoking during pregnancy on log-birthweight (linear quantile specification)
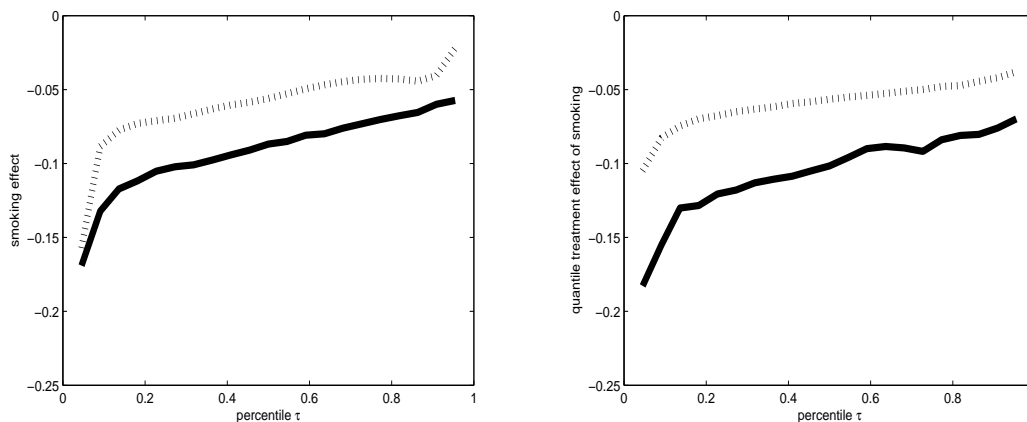
Note: Data from Abrevaya (2006). Left graph: solid line is the pooled quantile regression smoking coefficient; dashed line is the panel quantile regression smoking coefficient. Right graph: solid line is the raw quantile treatment effect of smoking; dashed line is the quantile treatment effect estimate based on panel quantile regression.

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. Estimates are computed using $L = 21$ knots. The stochastic EM algorithm is run for 100 iterations, with 100 random walk Metropolis-hastings draws within each iteration.[13] Parameter estimates are computed as averages of the 50 last iterations of the algorithm.

We see on the left graph of Figure 1 that the smoking effect becomes less negative when

---

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

Figure 2: Quantile effects of smoking during pregnancy on log-birthweight (interacted quantile specification)
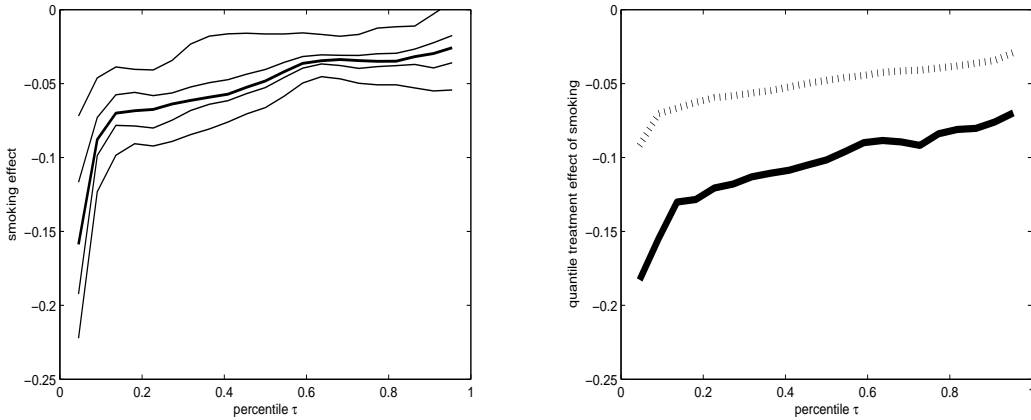
Note: Data from Abrevaya (2006). Left graph: lines represent the percentiles .05, .25, .50, .75, and .95 of the heterogeneous smoking effect across mothers, at various percentiles τ. Right graph: solid line is the raw quantile treatment effect of smoking; dashed line is the quantile treatment effect estimate based on panel quantile regression with interactions.

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 Subsection 2.3. This calculation illustrates the usefulness of specifying and estimating a complete semiparametric model of the joint distribution of outcomes and unobservables, in order to compute counterfactual distributions that take into account the presence of unobserved heterogeneity. On the graph, the solid line shows the empirical difference between unconditional quantiles, while the dashed line shows the quantile treatment effect that accounts for both observables and unobservables.

The results 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 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).

5   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 to flexibly model feedback processes in models with predetermined covariates. The empirical application illustrates the benefits of having a flexible approach to allow for heterogeneity and nonlinearity within the same model in a panel data context.

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 to deal 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 (2014) 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 (2014): "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] Cai, Z. and X. Xu (2008): "Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models," *Journal of the American Statistical Association*, 103, 1595–1608.

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

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

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

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

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

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

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

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

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

[24] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

[38] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

APPENDIX

A 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 $\lambda_L > 0$:

$$\theta(\tau) = \theta(\tau_1) + \frac{\ln(\tau/\tau_1)}{\lambda_1}\iota_c, \quad \tau \leq \tau_1,$$

$$\theta(\tau) = \theta(\tau_L) - \frac{\ln((1-\tau)/(1-\tau_L))}{\lambda_L}\iota_c, \quad \tau > \tau_L,$$

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:

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

Similarly, the approximating density of individual effects is:

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

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:

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

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 $f(\eta_i|W_i; \widehat{\xi}^{(s)})$.[14]

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:

$$\sum_{i=1}^{N} \int \Psi_i(\eta; \widehat{\xi}) f(\eta|W_i; \widehat{\xi}) d\eta = 0. \tag{B1}$$

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.[15]

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:

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

---

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

[15]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.

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

$$A\left(\widehat{\xi}^{(s+1)} - \widehat{\xi}\right) + B\left(\widehat{\xi}^{(s)} - \widehat{\xi}\right) + \varepsilon^{(s)} = o_p\left(N^{-\frac{1}{2}}\right), \tag{B2}$$

where:

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

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

and:

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

Note that:

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

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:

$$\sqrt{N}\left(\widehat{\xi}^{(s)} - \widehat{\xi}\right) = \sum_{k=0}^{\infty} \left(-A^{-1}B\right)^k \left(-A^{-1}\right) \sqrt{N}\varepsilon^{(s-1-k)} + o_p(1). \tag{B3}$$

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:

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

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 (31)-(32). 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 4 *(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 4.1. 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:

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

and

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

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:

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

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

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

Proposition B1. *(nonparametric consistency)*
Under Assumption B1, $\widehat{\xi}(\cdot)$ is uniformly consistent for $\overline{\xi}(\cdot)$ in the sense that (46) 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)*:

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

so, by Assumption B1 *(i)*, $\left\| \widetilde{\xi}(\cdot) - \overline{\xi}(\cdot) \right\|_\infty \leq \epsilon$ with probability approaching one. This shows (46).
∎

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 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:

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

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 $\mathbb{E}[Z_i Z_i']$ are bounded away from zero (as well as from above), then there exists a constant $\mu > 0$ such that:

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

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 model 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 2.4, 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:

$$A \equiv \sup_{\xi(\cdot)\in\mathcal{H}_K} \left|\frac{1}{L}\sum_{\ell=1}^L \left\|\frac{1}{N}\sum_{i=1}^N \varphi_i\left(\pi_L\xi(\cdot), \tau_\ell\right)\right\|^2 - \frac{1}{L}\sum_{\ell=1}^L \left\|\mathbb{E}\left[\varphi_i\left(\pi_L\xi(\cdot), \tau_\ell\right)\right]\right\|^2\right| = o_p(1),$$

$$B \equiv \sup_{\xi(\cdot)\in\mathcal{H}_K} \left|\frac{1}{L}\sum_{\ell=1}^L \left\|\mathbb{E}\left[\varphi_i\left(\pi_L\xi(\cdot), \tau_\ell\right)\right]\right\|^2 - \frac{1}{L}\sum_{\ell=1}^L \left\|\mathbb{E}\left[\varphi_i\left(\xi(\cdot), \tau_\ell\right)\right]\right\|^2\right| = o(1),$$

$$C \equiv \sup_{\xi(\cdot)\in\mathcal{H}_K} \left|\frac{1}{L}\sum_{\ell=1}^L \left\|\mathbb{E}\left[\varphi_i\left(\xi(\cdot), \tau_\ell\right)\right]\right\|^2 - \int_0^1 \left\|\mathbb{E}\left[\varphi_i\left(\xi(\cdot), \tau\right)\right]\right\|^2 d\tau\right| = o(1).$$

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 $(0,1)$:

$$\left\|\varphi_i\left(\xi_2(\cdot), \tau_2\right) - \varphi_i\left(\xi_1(\cdot), \tau_1\right)\right\| \leq C_1\sqrt{K}\left\|\xi_2(\cdot) - \xi_1(\cdot)\right\|_\infty^\nu + C_2\sqrt{K}|\tau_2 - \tau_1|. \tag{B4}$$

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

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

One possibility to establish (B4) is to assume that $\eta \mapsto W_{it}(\eta)'\theta(\tau)$ is invertible almost surely[16] and that its inverse is Lipschitz in $\theta(\tau)$, and then to use the expression of $f(\eta|Y_i, X_i; \pi_L \xi(\cdot))$, which involves the piecewise-linear expressions (47) and (48).

The $\pi_L \xi(\cdot)$ belong to a compact KL-dimensional space. Using (B4), 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 (B4) 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 (B4) but now for $\xi_1(\cdot) = \xi_2(\cdot)$, and using that $K/L^2 = o(1)$, yields $C = o(1)$.

C   Monte Carlo illustration

The simulated model is as follows:

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

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

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

where V_i is i.i.d. uniform on $(0, 1)$, independent of everything else, and where $\overline{X}_{1i}$ and $\overline{X}_{2i}$ denote individual averages. Lastly, $\beta_j(\tau)$ and $\delta_j(\tau)$ are defined on a set of $L = 11$ knots; see top row of Table D1. Here, as in the empirical analysis, we use linear splines to construct the approximating model. We also use the exponential modelling described in Appendix A for the extreme intervals. Lastly, we use the correct number of knots ($L = 11$).[17]

The top panel in Table D1 shows the estimates of $\beta_0(\tau_\ell)$, $\beta_1(\tau_\ell)$, $\beta_2(\tau_\ell)$, and $\gamma(\tau_\ell)$, across 100 simulated datasets with $N = 1000$ and $T = 3$. We report the population values of the parameters, and means and standard deviations across simulations. The results show moderate finite-sample biases, and relatively precise estimates, even at the extreme knots. We also observe larger biases for $\gamma(\cdot)$ and the constant coefficient. The bottom panel in Table D1 shows the estimates of $\delta_0(\tau_\ell)$, $\delta_1(\tau_\ell)$, and $\delta_2(\tau_\ell)$. In this case we observe somewhat larger standard errors in the tails. Nevertheless, biases throughout the distribution are moderate. Together, these results suggest reasonable finite-sample performance of our estimator.

D   Nonparametric identification: informal sketch of the arguments

We consider two setups in turn: the static model of Subsection 2.1, and the autoregressive first-order Markov model with exogenous regressors of Subsection 2.2. In both cases we provide a brief informal sketch of the identification argument.

---

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

[17]There are two small differences with the algorithm used in the empirical application in Section 4. Here we set $\lambda_1 = \lambda_1^\eta = 1 - \tau_1$ and $\lambda_L = \lambda_L^\eta = \tau_L$, and we do not estimate these parameters. In addition, we use a different, independent Metropolis-Hastings method to generate the η draws in the stochastic EM algorithm.

Static model. Consider the static model of Subsection 2.1. 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:

$$L_{Y_1,(y_2),Y_3} = L_{Y_3|\eta} \Delta_{(y_2)|\eta} L_{Y_1,\eta}, \quad y_2 - a.e. \tag{D5}$$

This yields a joint diagonalization system of operators, because, under suitable invertibility (i.e., injectivity) conditions, (D5) implies:

$$L_{Y_1,(y_2),Y_3} L_{Y_1,(\widetilde{y}_2),Y_3}^{-1} = L_{Y_3|\eta} \Delta_{(y_2)|\eta} \Delta_{(\widetilde{y}_2)|\eta}^{-1} L_{Y_3|\eta}^{-1}, \quad (y_2, \widetilde{y}_2) - a.e. \tag{D6}$$

The conditions of Hu and Schennach (2008)'s theorem then guarantee uniqueness of the solutions to (D6).

Dynamic autoregressive model. Let us now consider the dynamic autoregressive model of Subsection 2.2. 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:

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

43

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:

$$L_{Y_1,(y_2),(y_3),Y_4} = L_{Y_4|(y_3),\eta}\Delta_{(y_3)|(y_2),\eta}L_{Y_1,(y_2),\eta}, \quad (y_2,y_3) - a.e. \tag{D7}$$

Hence, under suitable invertibility conditions:

$$L_{Y_1,(y_2),(y_3),Y_4}L_{Y_1,(\widetilde{y}_2),(\widetilde{y}_3),Y_4}^{-1} = L_{Y_4|(y_3),\eta}\Delta_{(y_3)|(y_2),\eta}\Delta_{(\widetilde{y}_3)|(\widetilde{y}_2),\eta}^{-1}L_{Y_4|(\widetilde{y}_3),\eta}^{-1}, \quad (y_2,\widetilde{y}_2,y_3,\widetilde{y}_3) - a.e. \tag{D8}$$

Hu and Shum (2012), in particular in their Lemma 3, provide conditions for uniqueness of the solutions to (D8). Their conditions are closely related to the ones in Hu and Schennach (2008); see Assumption 4 in Subsection 2.4.

Table D1: Monte Carlo results

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

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