Empirical Economics (2021) 60:227–259
https://doi.org/10.1007/s00181-020-01893-5

Bayesian panel quantile regression for binary outcomes with correlated random effects: an application on crime recidivism in Canada

Georges Bresson[1] · Guy Lacroix[2] · Mohammad Arshad Rahman[3]

Received: 9 December 2019 / Accepted: 26 May 2020 / Published online: 23 June 2020

Abstract

This article develops a Bayesian approach for estimating panel quantile regression with binary outcomes in the presence of correlated random effects. We construct a working likelihood using an asymmetric Laplace error distribution and combine it with suitable prior distributions to obtain the complete joint posterior distribution. For posterior inference, we propose two Markov chain Monte Carlo (MCMC) algorithms but prefer the algorithm that exploits the blocking procedure to produce lower autocorrelation in the MCMC draws. We also explain how to use the MCMC draws to calculate the marginal effects, relative risk and odds ratio. The performance of our preferred algorithm is demonstrated in multiple simulation studies and shown to perform extremely well. Furthermore, we implement the proposed framework to study crime recidivism in Quebec, a Canadian Province, using novel data from administrative correctional files. Our results suggest that the recently implemented "tough-on-crime" policy of the Canadian government has been largely successful in reducing the probability of repeat offenses in the post-policy period. Besides, our results support existing findings on crime recidivism and offer new insights at various quantiles.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00181-020-01893-5) contains supplementary material, which is available to authorized users.

✉ Georges Bresson
georges.bresson@u-paris2.fr

Guy Lacroix
guy.lacroix@ecn.ulaval.ca

Mohammad Arshad Rahman
marshad@iitk.ac.in

1 Department of Economics, Université Paris II, 12 Place du Panthéon, 75231 Paris Cedex 05, France

2 Department of Economics, Université Laval, Québec, Canada

3 Department of Economic Sciences, Indian Institute of Technology Kanpur, Kanpur, India

Keywords Bayesian inference · Correlated random effects · Crime · Panel data · Quantile regression · Recidivism

JEL Classification C11 · C31 · C33 · C35 · K14 · K42

1 Introduction

The concept of quantile regression introduced in Koenker and Bassett (1978) has captured the attention of both statisticians and econometricians, theorists as well as applied researchers, and across school of thoughts i.e., Classicals (or Frequentists) and Bayesians. Quantile regression offers several advantages over mean regression such as robustness against outliers and desirable equivariance properties for monotone transformation. Estimation methods, particularly for cross-sectional data, are also well developed.[1] The method has been employed in various disciplines including economics, finance, and the social sciences (Koenker 2005; Davino et al. 2013). However, the development of quantile regression for panel data witnessed noticeable delay (more than two decades) because of complexities in estimation. The primary challenge was that quantiles, unlike means, are not linear operators and hence standard differencing (or demeaning) methods are not applicable for estimation of quantile regression. The challenges in estimation increases, if, for example, the outcome variable is discrete (such as binary or ordinal) because quantiles for such variables are not readily defined. Besides, modeling of panel data brings in consideration of unobserved individual-specific heterogeneity and the related debate on the choice of "random effects" versus "fixed effects." Motivated by these challenges in modeling and estimation, this paper considers a quantile regression model for panel data in the presence of correlated random effects (CRE) and introduces two Markov chain Monte Carlo (MCMC) algorithms for its estimation. The proposed framework is applied to study crime recidivism in the Province of Quebec, Canada, using novel data constructed from administrative correctional files.

The current paper touches on at least two growing econometric/statistic literatures—quantile regression for panel data and panel quantile regression for discrete outcomes. In reference to the former, Koenker (2004) introduced a penalization-based method for estimating panel quantile regression with fixed effects (i.e., the unobserved individual-specific effects are assumed to be fixed and from a heterogeneous population). Geraci and Bottai (2007) considered a panel quantile regression with individual-specific intercept and constructed a working likelihood using the asymmetric Laplace (AL) distribution (Yu and Moyeed 2001). They assumed a normal distribution on the individual-specific intercept and presented a Monte Carlo expectation–maximization (EM) algorithm for estimating the model. The proposed model and algorithm were

[1] Some Classical techniques include simplex method (Dantzig 1963; Dantzig and Thapa 1997, 2003; Barrodale and Roberts 1973; Koenker and d'Orey 1987), interior point algorithm (Karmarkar 1984; Mehrotra 1992) and smoothing algorithm (Madsen and Nielsen 1993; Chen 2007). Bayesian methods using Markov chain Monte Carlo (MCMC) algorithms for estimating quantile regression was introduced in Yu and Moyeed (2001) and refined, among others, in Kozumi and Kobayashi (2011). A non-Markovian simulation-based algorithm was proposed in Rahman (2013). See also Soares and Fagundes (2018) for interval quantile regression using swarm intelligence.

implemented to study labor pain data reported in Davis (1991). Liu and Bottai (2009) generalized the model of Geraci and Bottai (2007) by including multiple individual-specific effects and used a heavy-tailed distribution (multivariate Laplace distribution, instead of a multivariate normal distribution) for the multiple individual-specific effects. Later, Geraci and Bottai (2014) also considered the panel quantile regression model of Geraci and Bottai (2007) to accommodate multiple individual-specific effects and suggested strategies to reduce the computational burden of the Monte Carlo EM algorithm.[2] A Bayesian approach for estimating panel quantile regression was presented in Luo et al. (2012), where they propose a Gibbs sampling algorithm by exploiting the normal-exponential mixture representation of the AL distribution (Kozumi and Kobayashi 2011). Wang (2012) also utilized the AL density to develop a Bayesian estimation method for quantile regression in a parametric nonlinear mixed-effects model.

The latter of the papers mentioned in the previous paragraph have assumed that the unobserved individual-specific effects are random and uncorrelated with the regressors—also known as “random effects” in the classical econometrics literature. In contrast, when the individual-specific effects are assumed to be correlated with the regressors, the models have been termed as “fixed effects” model. Fixed effects models suffer from the limitation that they cannot estimate the coefficient for time-invariant regressors. So, when most of the variation in a regressor is located in the individual dimension (rather than in the time dimension), estimation of coefficients of time-varying regressors may be imprecise. Most disciplines in applied statistics, other than econometrics, use the random effects model (Cameron and Trivedi 2005). However, as shown in Baltagi (2013), most applied work in economics have settled the choice between the two specifications using the specification test proposed in Hausman (1978).

Between the questionable orthogonality assumption of the random effects model and the limitations of the fixed effects specification, lies the idea of correlated random effects (CRE). This concept is utilized in the current paper to soften the assertion of unobserved individual heterogeneity being uncorrelated with regressors. The CRE was introduced in Mundlak (1978), where he models the individual-specific effects as a linear function of the time averages of all the time-varying regressors. Hausman and Taylor (1981) proposed an alternative specification in which some of the time-varying and time-invariant regressors are related to the unobserved individual-specific effects.[3] Later, Chamberlain (1982, 1984) considered a richer model and defined the individual-specific effects as a weighted sum of the regressors. These CRE models lead to an estimator of the coefficients of the regressors that equals the fixed effects

[2] For other developments on panel data quantile regression see Lamarche (2010), Canay (2011), Chernozhukov et al. (2013), Galvao et al. (2013), Galvao and Kato (2017), Harding and Lamarche (2017), Graham et al. (2018), Galvao and Poirier (2019), and Gu and Volgushev (2019) to mention a few. We thank a referee for suggesting the last reference.

[3] Baltagi et al. (2003) suggested an alternative pretest estimator based on the Hausman–Taylor (HT) model. This pretest alternative considers an HT model in which some of the variables, but not all, may be correlated with the individual effects. The pretest estimator becomes the random effects estimator if the standard Hausman test is not rejected. The pretest estimator becomes the HT estimator if a second Hausman test (based on the difference between the FE and HT estimators) does not reject the choice of strictly exogenous regressors. Otherwise, the pretest estimator is the FE estimator.

estimator. The literature has numerous publications on the Hausman tests or the CRE models in a linear or nonlinear framework. We refer the reader to Arellano (1993), Wooldridge (2010), Baltagi (2013), Burda and Harding (2013), Greene (2015) and references therein. Most recently, Joshi and Wooldridge (2019) extended the CRE approach to linear panel data models when instrumental variables are needed and the panel is unbalanced.

Within the quantile regression for panel data literature, Abrevaya and Dahl (2008) incorporated the CRE to the quantile panel regression model and utilized it to study birth weight using balanced panel data from Arizona and Washington. They make certain simplifying assumptions which allows them to estimate the model using pooled linear quantile regression. Following the quantile regression framework of Abrevaya and Dahl (2008), Bache et al. (2013) considers a more restricted specification to model birth weight using an unbalanced panel data from Denmark. Arellano and Bonhomme (2016) introduced a class of QR estimators for short panels, where the conditional quantile response function of the unobserved heterogeneity is also specified as a function of observables. The literature on Bayesian panel quantile regression with CRE is limited to Kobayashi and Kozumi (2012), where they develop Bayesian quantile regression for censored dynamic panel data and proposed a Gibbs sampling algorithm to estimate the model. The initial condition problem arising due to the dynamic nature of the model was successfully managed using CRE. In addition, they implement the framework to study the persistence of medical expenditures using data from the RAND health insurance experiment.

The literature on panel quantile regression for discrete outcomes is quite sparse, and most of the work has only come recently.[4] Alhamzawi and Ali (2018) extended the Bayesian ordinal quantile regression, introduced in Rahman (2016), to panel data and use it to analyze treatment-related changes in illness severity using data from the National Institute of Mental Health Schizophrenia Collaborative (NIMHSC), previously analyzed in Gibbons and Hedeker (1993). Ghasemzadeh et al. (2018) proposed a Gibbs sampling algorithm to estimate Bayesian quantile regression for ordinal longitudinal response in the presence of nonignorable missingness and use it to analyze the Schizophrenia data of Gibbons and Hedeker (1993). Ghasemzadeh et al. (2020) developed a Bayesian quantile regression model for bivariate longitudinal mixed ordinal and continuous responses to study the relationship between reading ability and antisocial behavior among children using the Peabody Individual Achievement Test (PIAT) data. Most recently, Rahman and Vossmeyer (2019) considered a panel quantile regression model with binary outcomes and develop an efficient blocked sampling algorithm. They apply the framework to study female labor force participation and home ownership using data from the Panel Study of Income Dynamics (PSID).

This article contributes to the two literature by incorporating the CRE concept into the panel quantile regression model for binary outcomes. We present two MCMC algorithms—a simple (nonblocked) Gibbs sampling algorithm and another blocked Gibbs sampling algorithm that exploits the block sampling of parameters

[4] A body of work related to quantile regression for discrete outcomes include, but is not limited to, Kordas (2006), Benoit and Poel (2010), Alhamzawi (2016), Omata et al. (2017), Alhamzawi and Ali (2020) and Rahman and Karnawat (2019).

to reduce the autocorrelation in MCMC draws. We also explain how to calculate the marginal effects, relative risk and the odds ratio using the MCMC draws. The performance of the blocked algorithm is thoroughly tested in multiple simulation studies and shown to perform extremely well. Lastly, we implement the model to study crime recidivism in the Province of Quebec, Canada, using data from the administrative correction files for the period 2007–2017. The results provide strong support for including the CRE into the binary panel quantile regression framework. On the applied side, we find that the recently implemented "tough-on-crime" policy has been successful in reducing the probability of repeat offenses and this is most pronounced at the lower quantiles. Besides, our results confirm existing findings from recent studies on crime recidivism, such as, schooling (unemployment rate) is negatively (positively) associated with crime recidivism. Moreover, the marginal effects and relative risk show considerable variability across the considered quantiles.

The remainder of the paper is organized as follows. Section 2 introduces the binary panel regression model with CRE and the two MCMC algorithms for its estimation. Section 3 presents the simulation studies and discusses the performance of the algorithm. Section 4 discusses how to compute the marginal effects, relative risk and odds ratio using the MCMC draws. Section 5 implements the proposed framework to study crime recidivism in Quebec, a Canadian Province. Section 6 presents concluding remarks.

2 The model

We propose a binary quantile regression framework for panel data where the individual-specific effects are correlated with the covariates giving rise to CRE. The binary panel quantile regression with CRE (BPQRCRE) model can be conveniently expressed in the latent variable formulation of Albert and Chib (1993) as follows,

$$
\begin{aligned}
z_{\mathrm{it}} &= x'_{\mathrm{it}}\beta + \alpha_i + \varepsilon_{\mathrm{it}} \quad \forall\, i = 1, \ldots, n, \quad t = 1, \ldots, T_i, \\
y_{\mathrm{it}} &= \begin{cases} 1 & \text{if } z_{\mathrm{it}} > 0, \\ 0 & \text{otherwise,} \end{cases} \\
\alpha_i &\sim N(\overline{m}'_i \zeta, \sigma^2_\alpha),
\end{aligned}
\tag{1}
$$

where z_{it} is a continuous latent variable associated with the binary outcome y_{it}, $x'_{\mathrm{it}} = (x_{\mathrm{it},1}, x_{\mathrm{it},2}, \ldots, x_{\mathrm{it},k})$ is a $(1 \times k)$ vector of explanatory variables including the intercept, and β is the $(k \times 1)$ vector of common parameters. Note that the variable z is often interpreted as a latent utility differential between making versus not making a choice, such as between participating versus not participating in a labor market (Kordas 2006; Rahman and Vossmeyer 2019). The parameter α_i is the individual-specific effect assumed to be independently distributed as a normal distribution, i.e., $\alpha_i \sim N\left(\overline{m}'_i \zeta, \sigma^2_\alpha\right)$. Here $\overline{m}_{i,j} = \sum_{t=1}^{T_i} x_{\mathrm{it},j}/T_i$ (for $j = 2, \ldots, k$) and $\overline{m}'_i = (\overline{m}_{i,2}, \ldots, \overline{m}_{i,k})$ is a $(1 \times (k-1))$ vector of individual means of time-varying explanatory variables excluding the intercept. The dependence of α on the covariates (x) yields a CRE model (Mundlak 1978).[5] The error term ε_{it}, conditional on α_i, is assumed to be independently and identically distributed (iid) as an asymmetric Laplace (AL) distribution i.e., $\varepsilon_{\text{it}}|\alpha_i \overset{\text{iid}}{\sim} AL\,(0, 1, p)$, where p denotes the quantile. The AL error distribution is used to create a working likelihood and has been utilized in previous studies on longitudinal data models such as Luo et al. (2012) and Rahman and Vossmeyer (2019).[6]

In the proposed BPQRCRE framework, the modeling of individual-specific effects as a function of the means of the time-varying covariates is inspired from Mundlak (1978). Utilizing $\overline{m}_i'$ as a set of controls for unobserved heterogeneity is both intuitive and advantageous. It is intuitive because it estimates the effect of the covariates holding the time average fixed, and advantageous because it serves as a compromise between the questionable orthogonality assumptions of the random effects model and the limitation of the fixed effects specification which leads to the incidental parameters problem. The considered model reduces to the standard uncorrelated random effects case if we set $\zeta = 0$, i.e., assume α_i is independent of the covariates (Rahman and Vossmeyer 2019). Here, we note that Chamberlain (1982, 1984) allowed for correlation between α_i and the covariates x_{it}' (excluding the intercept) through a more general formulation: $\alpha_i \sim N\left(\sum_{t=1}^{T_i} x_{\text{it}}'\zeta_t, \sigma_\alpha^2\right)$. However, this approach is more involved for an unbalanced panel, particularly if endogenous attrition is the reason for the panel to be unbalanced (see Wooldridge 2010). Besides, the CRE specification has a number of virtues for nonlinear panel data models as underlined in Burda and Harding (2013) and Greene (2015). Hence, we prefer the approach presented in Mundlak (1978) compared to the method in Chamberlain (1980, 1982, 1984).

The BPQRCRE model as presented in Eq. (1) can be directly estimated using an MCMC algorithm, but the resulting posterior will not yield the full set of tractable conditional posteriors necessary for a Gibbs sampler. Therefore, as done in Luo et al. (2012) and Rahman and Vossmeyer (2019), we utilize the normal-exponential mixture representation of the AL distribution to facilitate Gibbs sampling (Kozumi and Kobayashi 2011). The mixture representation for ε_{it} can be written as follows,

$$\varepsilon_{\text{it}} = \theta w_{\text{it}} + \tau\sqrt{w_{\text{it}}}u_{\text{it}}, \tag{2}$$

where $u_{\text{it}} \sim N\,(0, 1)$ is mutually independent of $w_{\text{it}} \sim \mathscr{E}\,(1)$ with $\mathscr{E}$ representing the exponential distribution, and the constants are $\theta = \frac{1-2p}{p(1-p)}$ and $\tau^2 = \frac{2}{p(1-p)}$.

[5] In the Bayesian literature, the elements of β do not differ across individuals and are referred to as fixed effects, whereas the α_i's are referred to as random effects. This terminology differs from the one used in econometrics. In the latter, the α_i's are treated either as random variables, and hence referred to as random effects, or as constant but unknown parameters and thus referred to as fixed effects [see Greenberg (2012), Baltagi et al. (2018)].

[6] The quantile regression objective function appears in the exponent of the AL distribution. Therefore, the minimization of the quantile loss function is equivalent to the maximization of the log-likelihood from an AL distribution. There does not exist any other known distribution which has a one-to-one correspondence between the coefficients of classical quantile regression and Bayesian quantile regression. We thank a referee for suggesting that we be more explicit about the choice of the AL distribution.

The mixture representation gives access to the appealing properties of the normal distribution.

To implement the Bayesian approach, we stack the model across i. Define $z_i = (z_{i1}, \ldots, z_{iT_i})'$, $y_i = (y_{i1}, \ldots, y_{iT_i})'$, $X_i = (x'_{i1}, \ldots, x'_{iT_i})'$, $w_i = (w_{i1}, \ldots, w_{iT_i})'$, $D_{\tau\sqrt{w_i}} = \tau \text{ diag}(\sqrt{w_{i1}}, \ldots, \sqrt{w_{iT_i}})'$ and $u_i = (u_{i1}, \ldots, u_{iT_i})'$. The resulting hierarchical model can be written as,

$$
\begin{aligned}
& z_i = X_i\beta + \iota_{T_i}\alpha_i + w_i\theta + D_{\tau\sqrt{w_i}}u_i \quad \forall \ i = 1, \ldots, n, \\
& y_{\text{it}} = \begin{cases} 1 & \text{if } z_{\text{it}} > 0, \\ 0 & \text{otherwise}, \end{cases} \quad \forall \ i = 1, \ldots, n, ; \ t = 1, \ldots, T_i, \\
& \alpha_i \sim N\left(\overline{m}'_i\zeta, \sigma_\alpha^2\right) \quad w_{\text{it}} \sim \mathscr{E}(1), \quad u_{\text{it}} \sim N(0, 1), \\
& \beta \sim N_k(\beta_0, B_0) \quad \sigma_\alpha^2 \sim IG\left(\frac{c_1}{2}, \frac{d_1}{2}\right), \quad \zeta \sim N_{k-1}(\zeta_0, C_0),
\end{aligned} \tag{3}
$$

where ι_{T_i} is a $(T_i \times 1)$ vector of ones and the last line in Eq. (3) presents the prior distribution on the parameters. The notation $N_k(\cdot)$ denotes a multivariate normal distribution of dimension k, and $IG(\cdot)$ denotes an inverse-gamma distribution. We note that the form of the prior distribution on β holds a penalty interpretation on the quantile loss function (Koenker 2004). A normal prior on β implies an ℓ_2 penalty and has been used in Geraci and Bottai (2007), Yuan and Yin (2010), Luo et al. (2012) and Rahman and Vossmeyer (2019).

By Bayes' theorem, we express the "complete joint posterior" density as proportional to the product of complete likelihood function and the prior distributions as follows,

$$
\begin{aligned}
\pi(\beta, \alpha, z, w, \zeta, \sigma_\alpha^2 \mid y) &\propto \left\{ \prod_{i=1}^{n} f(y_i \mid z_i, \beta, \alpha_i, w_i, \zeta, \sigma_\alpha^2)\pi(z_i \mid \beta, \alpha_i, w_i, \zeta, \sigma_\alpha^2) \right. \\
&\quad \left. \times \pi(w_i)\pi(\alpha_i) \right\} \pi(\beta)\pi(\zeta)\pi(\sigma_\alpha^2) \\
&\propto \left\{ \prod_{i=1}^{n} \left[\prod_{t=1}^{T_i} f(y_{\text{it}} \mid z_{\text{it}}) \right] \pi(z_i \mid \beta, \alpha_i, w_i, \zeta, \sigma_\alpha^2)\pi(w_i)\pi(\alpha_i) \right\} \\
&\quad \times \pi(\beta)\pi(\zeta)\pi(\sigma_\alpha^2),
\end{aligned} \tag{4}
$$

where the first line assumes independence between prior distributions and second line follows from the fact that given z_{it}, the observed y_{it} is independent of all parameters because the second line of Eq. (3) determines y_{it} given z_{it} with probability 1. Substituting the distribution of the variables associated with the likelihood and the prior distributions in Eq. (4) yields the following expression,

$$
\begin{aligned}
\pi(\beta, \alpha, z, w, \zeta, \sigma_\alpha^2 \mid y) &\propto \left\{ \prod_{i=1}^{n} \prod_{t=1}^{T_i} \left[I(z_{\text{it}} > 0)I(y_{\text{it}} = 1) + I(z_{\text{it}} \leq 0)I(y_{\text{it}} = 0) \right] \right\} \\
&\times \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \left\{ (z_i - X_i\beta - \iota_{T_i}\alpha_i - w_i\theta)' D_{\tau\sqrt{w_i}}^{-2} (z_i - X_i\beta - \iota_{T_i}\alpha_i - w_i\theta) \right\} \right]
\end{aligned}
$$

$$\times \left\{ \prod_{i=1}^{n} |D^2_{\tau\sqrt{w_i}}|^{-\frac{1}{2}} \right\} \left(2\pi\sigma_\alpha^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_\alpha^2} \sum_{i=1}^{n} (\alpha_i - \overline{m}_i'\zeta)'(\alpha_i - \overline{m}_i'\zeta) \right]$$
$$\times \exp\left(-\sum_{i=1}^{n}\sum_{t=1}^{T_i} w_{\text{it}} \right) |B_0|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\beta - \beta_0)' B_0^{-1} (\beta - \beta_0) \right] |C_0|^{-\frac{1}{2}}$$
$$\times \exp\left[-\frac{1}{2} (\zeta - \zeta_0)' C_0^{-1} (\zeta - \zeta_0) \right] \times \left(\sigma_\alpha^2\right)^{-\left(\frac{c_1}{2}+1\right)} \exp\left[-\frac{d_1}{2\sigma_\alpha^2} \right]. \tag{5}$$

The complete joint posterior density in Eq. (5) does not have a tractable form, and thus simulation techniques are necessary for estimation. Similar to Rahman and Vossmeyer (2019), we adopt a Bayesian approach due to the following two reasons. First, the likelihood function of a discrete panel data model is analytically intractable which makes optimization difficult using standard hill-climbing techniques. Second, numerical simulation methods for discrete panel data models are often slow and difficult to implement as noted in Burda and Harding (2013) and others. Complete joint posterior distribution (Eq. 5) readily yields a full set of conditional distributions (outlined below) which can be employed to estimate the model using Gibbs sampling.

We can derive the conditional posteriors of the parameters and latent variables from joint posterior density (5) by a straightforward extension of the nonblocked sampling method presented in Rahman and Vossmeyer (2019). This is presented in Algorithm 1, and the derivations of the conditional posterior densities can be found in the supplementary material. The parameters β are sampled from an updated multivariate normal distribution. Similarly, the parameter α_i's are sampled from an updated multivariate normal distribution. The latent weight w is sampled element wise from a generalized inverse Gaussian (GIG) distribution (Devroye 2014). The variance σ_α^2 is sampled from an updated inverse-gamma (IG) distribution. The parameters ζ are sampled from an updated multivariate normal distribution. Last, the latent variable z is sampled element wise from an univariate truncated normal (TN) distribution. Note that while drawing each of the parameters or latent variables, we hold the remaining quantities fixed as presented in Algorithm 1.

The MCMC procedure presented in Algorithm 1 exhibits the conditional posterior distributions for the parameters and latent variables necessary for a Gibbs sampler. While this Gibbs sampler is straightforward, there is potential for poor mixing of the MCMC draws due to correlation between (β, α_i) and (z_i, α_i). This correlation arises because the variables corresponding to the parameters in α_i are often a subset of those in x_{it}'. Thus conditioning these items on one another leads to high autocorrelation in MCMC draws as demonstrated in Chib and Carlin (1999) and noted in Rahman and Vossmeyer (2019).

To avoid the high autocorrelation in MCMC draws, we present an alternative algorithm that jointly samples (β, z) in one block within the Gibbs sampler (see Rahman and Vossmeyer 2019, for more on the blocking procedure). The details of our blocked

Algorithm 1 Nonblocked sampling in the BPQRCRE model

1. Sample $\beta \mid \alpha, z, w \sim N_k\left(\widetilde{\beta}, \widetilde{B}\right)$ where,
$\widetilde{B}^{-1} = \left(\sum_{i=1}^{n} X_i' D_{\tau\sqrt{w_i}}^{-2} X_i + B_0^{-1}\right)$, and $\widetilde{\beta} = \widetilde{B}\left(\sum_{i=1}^{n} X_i' D_{\tau\sqrt{w_i}}^{-2}\left(z_i - \iota_{T_i}\alpha_i - w_i\theta\right) + B_0^{-1}\beta_0\right)$.
2. Sample $\alpha_i \mid \beta, z, w, \sigma_\alpha^2, \zeta \sim N\left(\widetilde{a}, \widetilde{A}\right)$ for $i = 1, \ldots, n$, where,
$\widetilde{A}^{-1} = \left(\iota_{T_i}' D_{\tau\sqrt{w_i}}^{-2}\iota_{T_i} + \sigma_\alpha^{-2}\right)$, and $\widetilde{a} = \widetilde{A}\left(\iota_{T_i}' D_{\tau\sqrt{w_i}}^{-2}(z_i - X_i\beta - w_i\theta) + \sigma_\alpha^{-2}\overline{m}_i'\zeta\right)$.
3. Sample $w_{\text{it}} \mid \beta, \alpha_i, z_{\text{it}} \sim GIG\left(\frac{1}{2}, \widetilde{\lambda}_{\text{it}}, \widetilde{\eta}\right)$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T_i$, where,
$\widetilde{\lambda}_{\text{it}} = \left(\frac{z_{\text{it}} - x_{\text{it}}'\beta - \alpha_i}{\tau}\right)^2$, and $\widetilde{\eta} = \left(\frac{\theta^2}{\tau^2} + 2\right)$.
4. Sample $\sigma_\alpha^2 \mid \alpha, \zeta \sim IG\left(\frac{\widetilde{c}_1}{2}, \frac{\widetilde{d}_1}{2}\right)$ where,
$\widetilde{c}_1 = (n + c_1)$, and $\widetilde{d}_1 = d_1 + \sum_{i=1}^{n}\left(\alpha_i - \overline{m}_i'\zeta\right)'\left(\alpha_i - \overline{m}_i'\zeta\right)$.
5. Sample $\zeta \mid \alpha, \sigma_\alpha^2 \sim N_{k-1}\left(\widetilde{\zeta}, \widetilde{\Sigma}_\zeta\right)$ where,
$\widetilde{\Sigma}_\zeta^{-1} = \left(\sigma_\alpha^{-2}\sum_{i=1}^{n}\overline{m}_i\overline{m}_i' + C_0^{-1}\right)$, and $\widetilde{\zeta} = \widetilde{\Sigma}_\zeta\left(\sigma_\alpha^{-2}\sum_{i=1}^{n}\overline{m}_i\alpha_i' + C_0^{-1}\zeta_0\right)$.
6. Sample the latent variable $z \mid \beta, \alpha, w$ for all values of $i = 1, \ldots, n$ and $t = 1, \ldots, T_i$ from an univariate truncated normal (TN) distribution as follows,
$$z_{\text{it}} \mid \beta, \alpha, w \sim \begin{cases} TN_{(-\infty,0]}\left(x_{\text{it}}'\beta + \alpha_i + w_{\text{it}}\theta, \tau^2 w_{\text{it}}\right) & \text{if } y_{\text{it}} = 0, \\ TN_{(0,\infty)}\left(x_{\text{it}}'\beta + \alpha_i + w_{\text{it}}\theta, \tau^2 w_{\text{it}}\right) & \text{if } y_{\text{it}} = 1. \end{cases}$$

sampler are described in Algorithm 2, and the derivations of the conditional posterior densities are presented in the supplementary material. Specifically, β is sampled marginally of α_i from a multivariate normal distribution. Then the latent variable z_i is sampled marginally of α_i from a truncated multivariate normal distribution denoted by $TMVN_{B_i}$, where B_i is the truncation region given by $B_i = \left(B_{i1} \times B_{i2} \times \cdots \times B_{iT_i}\right)$ such that B_{it} is the interval $(0, \infty)$ if $y_{\text{it}} = 1$ and the interval $(-\infty, 0]$ if $y_{\text{it}} = 0$. To draw from a truncated multivariate normal distribution, we utilize the method proposed in Geweke (1991, 2005) as implemented in Rahman and Vossmeyer (2019). This involves drawing from a series of conditional posteriors which are univariate truncated normal distributions. The parameter α_i's, conditional on $(\beta, z, w, \sigma_\alpha^2, \zeta)$, are sampled from an updated multivariate normal distribution. The latent weight w is sampled elementwise from a generalized inverse Gaussian (GIG) distribution (Devroye 2014). The variance σ_α^2 is sampled from an updated inverse-gamma (IG) distribution. Lastly, the parameters ζ are sampled from an updated multivariate normal distribution. Once again, while sampling each quantity of interest, we hold the remaining parameters or latent variables fixed as exhibited in Algorithm 2.

Algorithm 2 Blocked sampling in the BPQRCRE model

1. Sample (β, z_i) marginally of α in one block as follows.
   (a) Let $\Omega_i = \sigma_\alpha^2 J_{T_i} + D^2_{\tau\sqrt{w_i}}$ with $J_{T_i} = \iota_{T_i}\iota'_{T_i}$. Sample $\beta \mid z, w, \sigma_\alpha^2, \zeta \sim N_k\left(\widetilde{\beta}, \widetilde{B}\right)$ where,
   $$\widetilde{B}^{-1} = \left(\sum_{i=1}^{n} X_i'\Omega_i^{-1}X_i + B_0^{-1}\right), \text{ and } \widetilde{\beta} = \widetilde{B}\left(\sum_{i=1}^{n} X_i'\Omega_i^{-1}\left(z_i - \iota_{T_i}\overline{x}_i'\zeta - w_i\theta\right) + B_0^{-1}\beta_0\right).$$
   (b) Sample the vector $z_i \mid \beta, w_i, \sigma_\alpha^2, \zeta \sim TMVN_{B_i}\left(X_i\beta + \iota_{T_i}\overline{m}_i'\zeta + w_i\theta, \Omega_i\right)$ for all $i = 1, \ldots, n$, where $B_i = \left(B_{i1} \times B_{i2} \times \cdots \times B_{iT_i}\right)$ and B_{it} is the interval $(0, \infty)$ if $y_{\text{it}} = 1$ and the interval $(-\infty, 0]$ if $y_{\text{it}} = 0$. This is achieved by sampling z_i at the jth pass of the MCMC iteration using a series of conditional posterior distributions as follows:
   $$z_{\text{it}}^j \mid z_{i1}^j, \ldots, z_{i(t-1)}^j, z_{i(t+1)}^j, \ldots, z_{iT_i}^j \sim TN_{B_{\text{it}}}\left(\mu_{t|-t}, \Sigma_{t|-t}\right), \text{ for } t = 1, \ldots, T_i,$$
   where TN denotes a truncated normal distribution. The terms $\mu_{t|-t}$ and $\Sigma_{t|-t}$ are the conditional mean and variance, and are defined as,
   $$\mu_{t|-t} = x_{\text{it}}'\beta + \overline{m}_i'\zeta + w_{\text{it}}\theta + \Sigma_{t,-t}\Sigma_{-t,-t}^{-1}\left(z_{i,-t}^j - \left(X_i\beta + \iota_{T_i}\overline{x}_i'\zeta + w_i\theta\right)_{-t}\right),$$
   $$\Sigma_{t|-t} = \Sigma_{t,t} - \Sigma_{t,-t}\Sigma_{-t,-t}^{-1}\Sigma_{-t,t},$$
   where $z_{i,-t}^j = \left(z_{i1}^j, \ldots, z_{i(t-1)}^j, z_{i(t+1)}^{j-1}, \ldots, z_{iT_i}^{j-1}\right)'$, $\left(X_i\beta + \iota_{T_i}\overline{m}_i'\zeta + w_i\theta\right)_{-t}$ is a column vector with tth element removed, $\Sigma_{t,t}$ denotes the (t, t)th element of Ω_i, $\Sigma_{t,-t}$ denotes the tth row of Ω_i with element in the tth column removed and $\Sigma_{-t,-t}$ is the Ω_i matrix with tth row and tth column removed.
2. Sample $\alpha_i \mid \beta, z, w, \sigma_\alpha^2, \zeta \sim N\left(\widetilde{a}, \widetilde{A}\right)$ for $i = 1, \ldots, n$, where,
   $$\widetilde{A}^{-1} = \left(\iota_{T_i}' D^{-2}_{\tau\sqrt{w_i}}\iota_{T_i} + \sigma_\alpha^{-2}\right), \text{ and } \widetilde{a} = \widetilde{A}\left(\iota_{T_i}' D^{-2}_{\tau\sqrt{w_i}}\left(z_i - X_i\beta - w_i\theta\right) + \sigma_\alpha^{-2}\overline{m}_i'\zeta\right).$$
3. Sample $w_{\text{it}} \mid \beta, \alpha_i, z_{\text{it}} \sim GIG\left(\frac{1}{2}, \widetilde{\lambda}_{\text{it}}, \widetilde{\eta}\right)$ for $i = 1, \ldots, n$, and $t = 1, \ldots, T_i$, where,
   $$\widetilde{\lambda}_{\text{it}} = \left(\frac{z_{\text{it}} - x_{\text{it}}'\beta - \alpha_i}{\tau}\right)^2, \text{ and } \widetilde{\eta} = \left(\frac{\theta^2}{\tau^2} + 2\right).$$
4. Sample $\sigma_\alpha^2 \mid \alpha, \zeta \sim IG\left(\frac{\widetilde{c}_1}{2}, \frac{\widetilde{d}_1}{2}\right)$ where,
   $$\widetilde{c}_1 = (n + c_1), \text{ and } \widetilde{d}_1 = d_1 + \sum_{i=1}^{n}\left(\alpha_i - \overline{m}_i'\zeta\right)'\left(\alpha_i - \overline{m}_i'\zeta\right).$$
5. Sample $\zeta \mid \alpha, \sigma_\alpha^2 \sim N_{k-1}\left(\widetilde{\zeta}, \widetilde{\Sigma}_\zeta\right)$ where,
   $$\widetilde{\Sigma}_\zeta^{-1} = \left(\sigma_\alpha^{-2}\sum_{i=1}^{n}\overline{m}_i\overline{m}_i' + C_0^{-1}\right), \text{ and } \widetilde{\zeta} = \widetilde{\Sigma}_\zeta\left(\sigma_\alpha^{-2}\sum_{i=1}^{n}\overline{m}_i\alpha_i' + C_0^{-1}\zeta_0\right).$$

3 A Monte Carlo simulation study

In this section, we present two simulation studies to demonstrate the performance of the blocked algorithm for the BPQRCRE model. The simulation data are generated from the following model,

$$\begin{aligned} z_{\text{it}} &= x_{\text{it}}'\beta + \alpha_i + \varepsilon_{\text{it}}, \quad \forall\, i = 1, \ldots, n, \quad \text{and} \quad t = 1, \ldots, T_i, \\ \alpha_i &= \overline{m}_i'\zeta + \xi_i, \quad \xi_i \sim N\left(0, \sigma_\alpha^2\right). \end{aligned} \tag{6}$$

where $x_{\text{it}}' = \left[1, x_{\text{it},2}, x_{\text{it},3}, x_{\text{it},4}\right]$, $\overline{m}_i' = \left[\overline{m}_{i,3}, \overline{m}_{i,4}\right]$, $\overline{m}_{i,j} = \sum_{t=1}^{T_i} x_{\text{it},j}/T_i$, $j = 3, 4$, $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)' = (0.5, 1, 0.6, -0.8)'$, $\zeta = (\zeta_3, \zeta_4)' = (-1, 1)'$. The covariates are generated as $x_{\text{it},2} \sim U(-2, 2)$, $x_{\text{it},3} \sim U(-2, 2)$, $x_{\text{it},4} \sim U(-2, 2)$, where U denotes a uniform distribution, and $\sigma_\alpha^2 = 1$. Our first sample is unbalanced with $n = 1000$ and $T_i \sim U(5, 15)$, leading to $T = \sum_{i=1}^{n} T_i = 9989$ observations. In a second exercise, we increase the number of individuals $n = 2000$ leading to $T = 19,985$ observations. The error term is generated from a standard AL distribution, i.e., $\varepsilon_{\text{it}} \sim AL(0, 1, p)$ for $i = 1, \ldots, n$, and $t = 1, \ldots, T_i$ at three different quantiles $p = 0.25, 0.5, 0.75$.

The binary outcome variable y is constructed from the continuous variable z, by assigning $y_{\text{it}} = 1$ whenever $z_{\text{it}} > 0$ and $y_{\text{it}} = 0$ whenever $z_{\text{it}} \leq 0$ for all of $i = 1, \ldots, n$ and $t = 1, \ldots, T_i$. We note that the binary response values of 0s and 1s are different at each quantile, because the error values generated from an AL distribution are different for each quantile. In the first simulation exercise with $n = 1000$, the number of observations corresponding to 0s and 1s for the 25th, 50th and 75th quantiles is (2283, 7706), (4217, 5772) and (6442, 3547), respectively. In the second simulation exercise with $n = 2000$, the number of observations corresponding to 0s and 1s for the 25th, 50th and 75th quantiles is (4640, 15345), (8691, 11294) and (13234, 6751), respectively. To complete the Bayesian setup for estimation, we use the following independent prior distributions: $\beta \sim N_k\left(0_k, 10^3 I_k\right)$, $\zeta \sim N_{k-2}\left(0_{k-2}, 10^3 I_{k-2}\right)$, $\sigma_\alpha^2 \sim IG\,(10/2, 9/2)$. For each exercise, we generate 16, 000 MCMC samples where the first 1000 values are discarded as burn-ins. The posterior estimates are reported based on the remaining 15, 000 MCMC iterations with a thinning factor of 10. The mixing of the MCMC chain is extremely good as illustrated in Fig. 1, which exhibits the trace and autocorrelation plots of the parameters from the second simulation exercise at the 75th quantile. The figure shows that, as desired, the chains mix well and the autocorrelation of the MCMC draws is close to zero. The plots from the first simulation exercise and the remaining quantiles in the second simulation exercise are extremely similar and not presented to avoid repetition and keep the paper within reasonable length. To supplement the plots in Fig. 1, Table 1 presents the autocorrelation in MCMC draws at lag 1, lag 5, and lag 10 confirming the good mixing across simulation exercises and at all quantiles.

The results from the two simulation exercises are presented in Table 2. Specifically, the table reports the true values of the parameters used to generate the data, along with the posterior mean, standard deviation and inefficiency factor (calculated using the batch-means method discussed in Greenberg 2012) of the MCMC draws. In general, the results show that the posterior means for (β, ζ) are near to their respective true values, $\beta = (0.5, 1, 0.6, -0.8)'$ and $\zeta = (-1, 1)'$ across all considered quantiles. The posterior standard deviations for all the parameters are small, and all the coefficients are statistically different from zero. So, the proposed MCMC algorithm is successful in correctly estimating all the model parameters across all quantiles. This is especially important because the number of 0s and 1s was different for each quantile. Moreover, the inefficiency factor for all the parameters is close to 1, suggesting a good sampling performance and a nice mixing of the Markov chain. Comparing the results from the first and second simulation exercise, we see that when the sample size is increased from $(n = 1000, T = 9989)$ to $(n = 2000, T = 19,985)$, the results improve and the posterior means of the coefficients are closer to their true values. In particular, some small observed biases for β_1, ζ_3, and ζ_4 at the 25th quantile are reduced to a large extent.

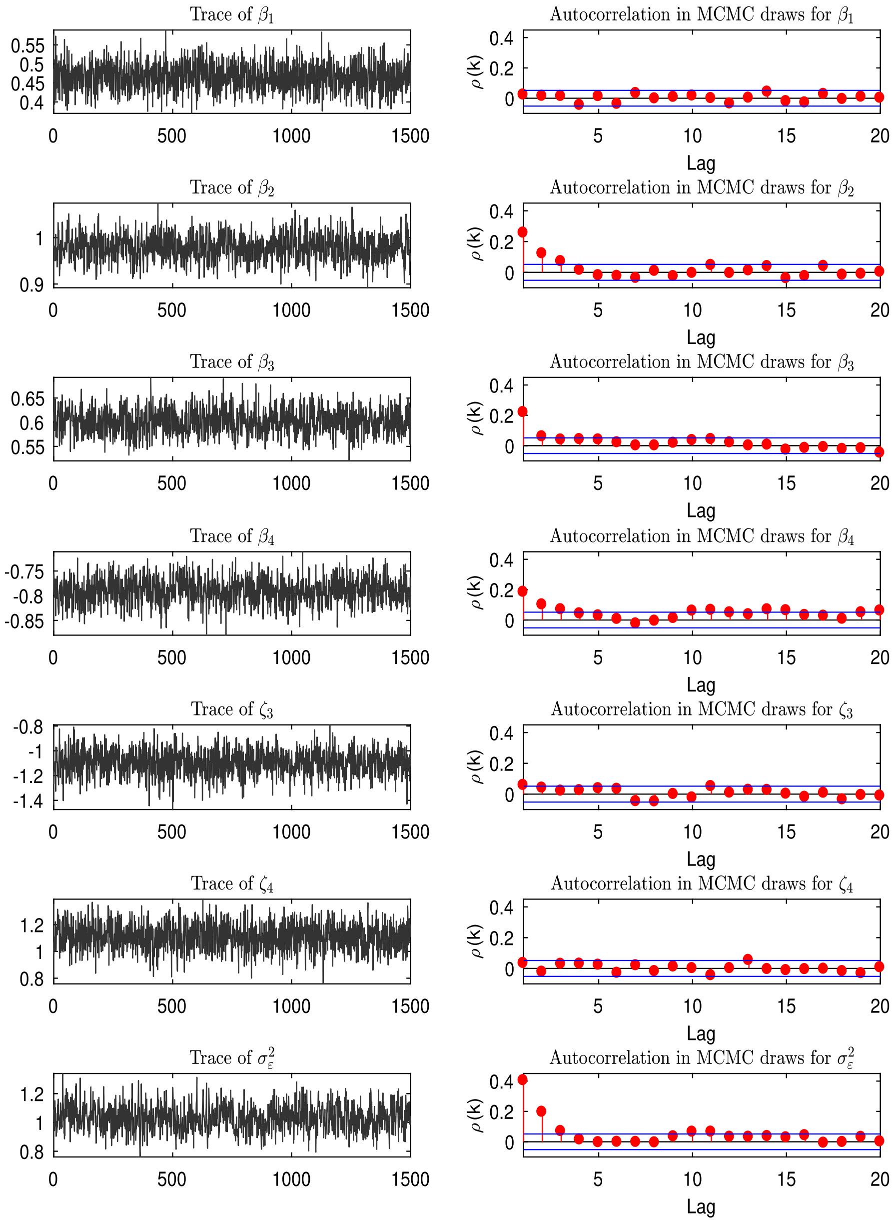

Fig. 1 Trace plots and autocorrelation plots of the parameters for the 75th quantile and $n = 2000$ individuals

Table 1 Autocorrelation in MCMC draws at lag 1, lag 5, and lag 10 for $n = 1000$ individuals (upper panel) and $n = 2000$ individuals (lower panel)

	25th quantile			50th quantile			75th quantile		
	Lag 1	Lag 5	Lag 10	Lag 1	Lag 5	Lag 10	Lag 1	Lag 5	Lag 10
$n = 1000$									
β_1	0.1351	0.0338	– 0.0258	0.0544	– 0.0079	– 0.0382	– 0.0417	– 0.0652	0.0165
β_2	0.3066	0.0369	0.0161	0.2385	0.0099	– 0.0218	0.2688	– 0.0567	0.0253
β_3	0.2828	0.0730	– 0.0003	0.1745	0.0012	– 0.0228	0.1784	– 0.0215	– 0.0125
β_4	0.3372	0.0783	0.0179	0.2421	0.0037	0.0348	0.1871	– 0.0254	– 0.0617
ζ_3	0.0653	0.0160	– 0.0314	0.0389	– 0.0080	0.0034	0.0669	– 0.0388	– 0.0338
ζ_4	0.1438	0.0319	– 0.0252	0.0649	– 0.0217	– 0.0721	0.0793	– 0.0317	0.0362
σ_α^2	0.4439	0.0658	0.0274	0.3115	– 0.0004	– 0.0181	0.3122	0.0151	– 0.0050
$n = 2000$									
β_1	0.1353	0.0200	– 0.0296	0.0176	0.0207	0.0154	0.0200	0.0096	0.0134
β_2	0.3092	0.0035	0.0189	0.3022	– 0.0151	– 0.0640	0.2539	– 0.0229	– 0.0079
β_3	0.1679	0.0655	0.0404	0.2051	– 0.0201	– 0.0142	0.2171	0.0367	0.0325
β_4	0.2648	0.0359	0.0222	0.2634	0.0415	– 0.0073	0.1816	0.0262	0.0575
ζ_3	0.0328	– 0.0098	0.0132	0.0762	– 0.0567	0.0017	0.0553	0.0340	– 0.0261
ζ_4	0.0782	– 0.0415	– 0.0178	0.0117	0.0179	0.0137	0.0314	0.0198	– 0.0022
σ_α^2	0.4381	0.0423	0.0227	0.3139	– 0.0189	0.0215	0.4017	– 0.0072	0.0621

We can allow the coefficients of x'_{it} and of $\overline{m}'_i$ to vary across the quantiles.[7] Let F_v denote the CDF of v_{it}. The pth quantile coefficients of x'_{it} are: $\beta(p) = \beta + 0.5F_v^{-1}(p)$ and the pth quantile coefficients of $\overline{m}'_i$ are: $\zeta(p) = \zeta + 0.3F_v^{-1}(p)$. We assume v_{it} follows a $N(0, 1)$ distribution.[8] For the sake of brevity, we only consider the larger sample ($n = 2000$, $T = 19{,}985$). The number of 0s and 1s for the 25th, 50th, and 75th quantiles is (5288, 14,697), (8691, 11,294), and (12,454, 7531), respectively. According to the estimation results in Table 3, the posterior means for $(\beta(p), \zeta(p))$ are near to their respective quantile-dependent true values across the three quantiles. In addition, the posterior standard deviations of all the parameters are small and all the coefficients are statistically different from zero. Finally, the small inefficiency factors indicate that the Markov chains mix well. It follows that the proposed algorithm for estimating BPQRCRE models performs well both when considering quantile-independent and quantile-dependent parameters, though as always, it performs better with larger samples. In that sense, the algorithm may be said to accommodate a large class of models.

[7] We thank a referee for this suggestion.

[8] For the 3 quantiles $p = 0.25, 0.5, 0.75$, we have $F_v^{-1}(0.25) = -0.6745$, $F_v^{-1}(0.5) = 0$ and $F_v^{-1}(0.75) = 0.6745$.

Table 2 True values (True), posterior mean (Mean), standard deviation (SD), and inefficiency factor (IF) of the parameters in the simulation study

	TRUE	25th quantile			50th quantile			75th quantile		
		MEAN	SD	IF	MEAN	SD	IF	MEAN	SD	IF
$n = 1000$										
β_1	0.5	0.7155	0.0582	1.2290	0.5799	0.0480	1.0544	0.5319	0.0523	0.9583
β_2	1.0	1.0093	0.0437	1.7885	0.9355	0.0331	1.3766	1.0155	0.0383	1.5226
β_3	0.6	0.7284	0.0403	1.5750	0.5898	0.0310	1.2686	0.5616	0.0372	1.2588
β_4	− 0.8	− 0.8699	0.0432	1.8581	− 0.7587	0.0330	1.3724	− 0.8482	0.0369	1.2620
ζ_3	− 1.0	− 1.2082	0.1493	1.0653	− 1.2043	0.1304	1.0389	− 1.0786	0.1451	1.0669
ζ_4	1.0	1.2781	0.1548	1.1919	1.0350	0.1373	1.0649	1.1079	0.1427	1.0793
σ_α^2	1.0	1.1668	0.1502	2.1466	1.1444	0.1177	1.6006	1.1923	0.1321	1.6553
$n = 2000$										
β_1	0.5	0.5241	0.0375	1.2201	0.4812	0.0326	1.0176	0.4661	0.0355	1.0200
β_2	1.0	0.9852	0.0281	1.6192	0.9985	0.0249	1.6350	0.9784	0.0274	1.5347
β_3	0.6	0.6134	0.0262	1.2643	0.5914	0.0226	1.3154	0.6017	0.0259	1.3277
β_4	− 0.8	− 0.7745	0.0278	1.4142	− 0.7719	0.0235	1.4121	− 0.7897	0.0253	1.3079
ζ_3	− 1.0	− 0.9418	0.0970	1.0328	− 1.0325	0.0894	1.0762	− 1.0957	0.1005	1.0553
ζ_4	1.0	0.9678	0.0985	1.0782	1.0814	0.0913	1.0117	1.1127	0.0994	1.0314
σ_α^2	1.0	0.8584	0.0857	2.1350	0.9433	0.0754	1.6048	1.0303	0.0895	1.8290

The upper panel presents results for $n = 1000$ individuals and the lower panel presents results for $n = 2000$ individuals

Table 3 True values (True), posterior mean (Mean), standard deviation (SD), and inefficiency factor (IF) of the quantile dependent parameters in the simulation study for $n = 2000$ individuals

	25th quantile				50th quantile				75th quantile			
	TRUE	MEAN	SD	IF	TRUE	MEAN	SD	IF	TRUE	MEAN	SD	IF
$\beta_1(p)$	0.163	0.182	0.035	0.997	0.5	0.481	0.032	1.017	0.837	0.787	0.037	1.001
$\beta_2(p)$	0.663	0.641	0.025	1.263	1	0.998	0.024	1.635	1.337	1.331	0.030	1.769
$\beta_3(p)$	0.263	0.269	0.024	1.080	0.6	0.591	0.022	1.315	0.937	0.939	0.029	1.521
$\beta_4(p)$	−1.137	−1.087	0.029	1.740	−0.8	−0.772	0.023	1.412	−0.462	−0.458	0.026	1.292
$\zeta_3(p)$	−1.202	−1.140	0.094	1.082	−1.0	−1.032	0.089	1.076	−0.797	−0.976	0.102	1.059
$\zeta_4(p)$	0.798	0.757	0.092	1.094	1	1.081	0.091	1.011	1.202	1.324	0.104	1.067
σ_α^2	1	0.837	0.078	1.810	1	0.943	0.075	1.604	1	1.025	0.092	1.972

4 Marginal effects, relative risk, and odds ratio

Our proposed binary panel quantile model is nonlinear, as such the coefficients by themselves do not give the marginal effects (Rahman 2016; Rahman and Vossmeyer 2019). However, marginal effects are important to understand the effect of a covariate on the probability of success. For example, in our current application one may be interested in seeing how the probability of recidivism is affected due to an additional year of schooling, decreasing regional unemployment rate by 1 percentage, or involvement in violent crime. These may be useful to policy makers and researchers alike.

To formally derive the marginal effects, we rewrite the BPQRCRE model presented in Eq. (1) as follows,

$$
\begin{aligned}
& z_{\mathrm{it}} = x_{\mathrm{it}}'\beta + \alpha_i + \varepsilon_{\mathrm{it}}, \quad \forall\, i = 1, \ldots, n, \quad \text{and} \quad t = 1, \ldots, T_i, \\
& \alpha_i \sim N(\overline{m}_i'\zeta, \sigma_\alpha^2),
\end{aligned}
\tag{7}
$$

where $\varepsilon_{\mathrm{it}} = w_{\mathrm{it}}\theta + \tau\sqrt{w_{\mathrm{it}}}u_{\mathrm{it}}$. We know $\varepsilon_{\mathrm{it}} \overset{\text{iid}}{\sim} AL(0, 1, p)$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T_i$, which implies $z_{\mathrm{it}}|\alpha_i \overset{\text{ind}}{\sim} AL(x_{\mathrm{it}}'\beta + \alpha_i, 1, p)$, where ind denotes independently distributed.

Given the model framework, the probability of success can be calculated as,

$$
\begin{aligned}
\Pr(y_{\mathrm{it}} = 1|x_{\mathrm{it}}, \beta, \alpha_i) &= \Pr(z_{\mathrm{it}} > 0|\beta, \alpha_i, x_{\mathrm{it}}) \\
&= 1 - \Pr(z_{\mathrm{it}} \le 0|\beta, \alpha_i, x_{\mathrm{it}}) \\
&= 1 - \Pr(\varepsilon_{\mathrm{it}} \le -x_{\mathrm{it}}'\beta - \alpha_i|\beta, \alpha_i, x_{\mathrm{it}}) \\
&= 1 - F_{\mathrm{AL}}(-x_{\mathrm{it}}'\beta - \alpha_i, 0, 1, p),
\end{aligned}
\tag{8}
$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T_i$, where $F_{\mathrm{AL}}(x, 0, 1, p)$ denotes the cumulative distribution function (cdf) of an AL distribution evaluated at x, with location 0, scale 1 and quantile p.

Marginal effect (i.e., the derivative of the probability of success with respect to a covariate) is often computed at the average covariate values or by averaging the marginal effects over the sample, *alias* average partial effects (Wooldridge 2010; Greene 2017). However, Jeliazkov and Vossmeyer (2018) show that both these quantities can be clearly inadequate in nonlinear settings (e.g., binary, ordinal and Poisson models) because they employ point estimates rather than their full distribution. To account for the uncertainty in parameters, we need another layer of integration over the model parameters. This idea of calculating the marginal effect that accounts for uncertainty in parameters and the covariates has been previously considered, among others, by Chib and Jeliazkov (2006) in the context of semiparametric dynamic binary longitudinal models, and Jeliazkov et al. (2008) and Jeliazkov and Rahman (2012) in relation to ordinal and binary models. Within the quantile literature, this has been mentioned by Rahman (2016) in the context of ordinal models and discussed by Rahman and Vossmeyer (2019) in connection to binary longitudinal outcome models.

Suppose, we are interested in the average marginal effect, i.e., average difference between probabilities of success when the jth covariate $\{x_{\mathrm{it},j}\}_{t=1}^{T_i}$ is set to the values

a and b, denoted as $\{x^a_{\text{it},j}\}_{t=1}^{T_i}$ and $\{x^b_{\text{it},j}\}_{t=1}^{T_i}$, respectively. To proceed, we split the covariate and parameter vectors as follows: $x^a_{\text{it}} = (x^a_{\text{it},j}, x_{\text{it},-j})$, $x^b_{\text{it}} = (x^b_{\text{it},j}, x_{\text{it},-j})$, and $\beta = (\beta_j, \beta_{-j})$, where $-j$ in the subscript denotes all covariates/parameters except the jth covariate/parameter. We are interested in the distribution of the difference $\{\Pr(y_{\text{it}} = 1|x^b_{\text{it},j}) - \Pr(y_{\text{it}} = 1|x^a_{\text{it},j})\}$, marginalized over $\{x_{\text{it},-j}\}$ and (β, α), given the data $y = (y_1, \ldots, y_n)'$. As done in Chib and Jeliazkov (2006) and Rahman and Vossmeyer (2019), we marginalize the covariates using their empirical distribution and integrate the parameters using their posterior distribution.

To obtain a sample of draws from the distribution of the difference in probabilities of success, marginalized over $(\{x_{\text{it},-j}\}, \beta, \alpha)$, given the data $y = (y_1, \ldots, y_n)'$, we express the distribution as follows,

$$
\begin{aligned}
&\{\Pr(y_{\text{it}} = 1|x^b_{\text{it},j}) - \Pr(y_{\text{it}} = 1|x^a_{\text{it},j})\} \\
&\quad = \int \Big\{ P(y_{\text{it}} = 1|x^b_{\text{it},j}, x_{\text{it},-j}, \beta, \alpha) - P(y_{\text{it}} = 1|x^a_{\text{it},j}, x_{\text{it},-j}, \beta, \alpha) \Big\} \\
&\quad \times \pi(x_{\text{it},-j}) \pi(\beta|y) \pi(\alpha|y) \, d(x_{\text{it},-j}) \, d\beta \, d\alpha.
\end{aligned} \tag{9}
$$

Drawing a sample from the above predictive distribution (i.e., Eq. 9) utilizes the method of composition. This involves randomly drawing an individual, extracting the corresponding sequence of covariate values, drawing a value (β, α) from the posterior distribution, and finally evaluating $\{\Pr(y_{\text{it}} = 1|x^b_{\text{it},j}, x_{\text{it},-j}, \beta, \alpha) - \Pr(y_{\text{it}} = 1|x^a_{\text{it},j}, x_{\text{it},-j}, \beta, \alpha)\}$. This is repeated for all other individuals and other draws from the posterior distribution. Finally, the average marginal effect ($\text{AME}_{\text{Bayes}}$) is calculated as the average of the difference in pointwise probabilities of success as follows,

$$
\begin{aligned}
\text{AME}_{\text{Bayes}} \approx \frac{1}{T}\frac{1}{M} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \sum_{m=1}^{M} \Big[& F_{\text{AL}}(-x^a_{\text{it},j}\beta_j^{(m)} - x'_{\text{it},-j}\beta_{-j}^{(m)} - \alpha_i^m, 0, 1, p) \\
& - F_{\text{AL}}(-x^b_{\text{it},j}\beta_j^{(m)} - x'_{\text{it},-j}\beta_{-j}^{(m)} - \alpha_i^m, 0, 1, p) \Big]
\end{aligned} \tag{10}
$$

where the expression for probability of success follows from Eq. (8), $T = \sum_{i=1}^{n} T_i$ is the total number of observations, and M is the number of MCMC draws. Here, $(\beta^{(m)}, \alpha^{(m)})$ is an MCMC draw of (β, α) for $m = 1, \ldots, M$. The quantity in Eq. (10) provides estimate that integrates out the variability in the sample and the uncertainty in parameter estimation.

Relative risk (RR) can be calculated to demonstrate the association between the risk factor or exposure (x_j) and the event (y) being studied. It is the ratio of the probability of the outcome with the risk factor $(x_j = b)$ to the probability of the outcome with the risk factor $(x_j = a)$ (e.g., exposed $(b = 1)$ /nonexposed $(a = 0)$). Following Eq. (10), the relative risk can be computed as,

$$
\text{RR}(b/a)_{\text{Bayes}} = \frac{1}{T}\frac{1}{M} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \sum_{m=1}^{M} \frac{H^b_{\text{AL}}}{H^a_{\text{AL}}}. \tag{11}
$$

where $H_{\mathrm{AL}}^{r} = 1 - F_{\mathrm{AL}}(-x_{\mathrm{it},j}^{r}\beta_j^{(m)} - x_{\mathrm{it},-j}^{\prime}\beta_{-j}^{(m)} - \alpha_i^m, 0, 1, p)$ for $r = a, b$, is the complement of the cdf of the AL distribution. If there is a causal effect between the exposure and the outcome, values of RR can be interpreted as follows: if RR $>$ 1 (resp. RR $<$ 1), the risk of outcome is increased (resp. decreased) by the exposure, and if RR = 1, the exposure does not affect the outcome.

The odds ratio is the ratio of the odds of the event occurring with the risk factor ($x_j = b$) to the odds of it occurring with the risk factor ($x_j = a$). It can be computed as:

$$\mathrm{OR}(b/a)_{\mathrm{Bayes}} = \frac{1}{T}\frac{1}{M}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\sum_{m=1}^{M}\left(\frac{H_{\mathrm{AL}}^{b}}{1 - H_{\mathrm{AL}}^{b}}\right) \Big/ \left(\frac{H_{\mathrm{AL}}^{a}}{1 - H_{\mathrm{AL}}^{a}}\right). \qquad (12)$$

The odds ratio, for a given exposure x_j, does not have an intuitive interpretation as the relative risk. OR is often interpreted as if they were equivalent to relative risks while ignoring their meaning as a ratio of odds. Two main factors influence the discrepancies between RR and OR: the initial risk of an event y_{it}, and the strength of the association between exposure $x_{\mathrm{it},j}$ and the event y_{it}. When the event $y_{\mathrm{it}} = 1$ is rare, then $\mathrm{OR}(b/a) \approx \mathrm{RR}(b/a)$, but the odds ratio generally overestimates the relative risk, and this overestimation becomes larger with increasing incidence of the outcome.

5 An application to crime recidivism in Canada

Crime has been extensively studied by economists both theoretically and empirically (see, e.g., Chalfin and McCrary (2017) for a recent survey). Many empirical analyses have used panel data either at the state (Cornwell and Trumbull 1994; Baltagi 2006; Baltagi et al. 2018) or at the individual level (Bhuller et al. 2020). The vast majority of the published papers focus on the situation in the USA. Here, we study crime recidivism in Canada between 2007 and 2017 for two reasons. First, the Canadian government implemented a "tough-on-crime" policy in 2012 which marked a shift from rehabilitating to warehousing people. Our proposed estimator is well suited to measure the sensitivity of recidivism to this new policy.[9] Second, offenders who are sentenced to less than 2 years serve their sentence in a provincial *correctional institution*, while offenders sentenced to 2 years or more serve their's in a *federal penitentiary*. The former have committed less serious crimes and are more likely to reoffend over the time span of our panel. Because our analysis focuses on this population, the impact of the "tough-on-crime" policy may be more easily unearthed from the data than if it focused on detainees serving long sentences.

5.1 The data

We use a sample drawn from the administrative correctional files for the Province of Quebec. The files are used by corrections personnel to manage activities and inter-

[9] Starting in 2012, the government enacted a series of legislations that made prison conditions more austere; imposed lengthier incarceration periods; significantly expanded the scope of mandatory minimum penalties; and reduced opportunities for conditional release, parole, and alternatives to incarceration.

Table 4 Descriptive summary of the sample data

	Mean	SD
Age	41.366	12.596
Schooling	6.011	3.814
Married	0.045	0.208
Aboriginal†	0.045	0.206
Mother tongue not Fr. or Eng.	0.070	0.255
Type of crime		
Traffic related	0.163	0.384
Violent (Domestic, assault & battery, etc.)	0.099	0.299
Property (Theft, robbery, etc.)	0.439	0.496
Other infractions to criminal code	0.299	0.458
Unemployment rate	8.329	2.063
Post 2012 (= 1)	0.741	0.437
Recidivism entire sample	0.114	0.318
Recidivism early entrants (Before 2012)	0.091	0.288
Recidivism late entrants (After 2012)	0.023	0.150

† First Nations, Inuit and Métis

ventions related to housing offenders and contain detailed information on inmates' characteristics, correctional facilities, and sentence administration. While they offer a wealth of information, the files have never been used for research purposes. We have drawn a random sample of 8974 detainees out of a population of 148,441. Each detainee is observed upon release and up until 2017. The earliest releases occur in 2007 and the latest in 2016. Overall, our unbalanced panel includes 61,880 observations. Of the 8974 detainees, as many as 3466 had at least one repeat offense over our sample period.

Table 4 presents the main characteristics of our sample. Detainees are 41 years of age on average, have a level of schooling corresponding to a high-school degree, and few are married. Aboriginal detainees represent 4.5% of our sample, and most are incarcerated in a correctional institution suited to their needs and specificities. Approximately 7% of inmates do not have French or English, Canada's two official languages, as their mother tongue. These include some Aboriginal residents as well as recent immigrants. Crimes have been aggregated into 4 distinct categories. By far the most common concerns property crime. Traffic related and infractions to the criminal code usually entail shorter sentences. Violent crimes receive the longest sentences in our data but necessarily less than 2 years. As mentioned above, major crimes fall under the federal jurisdiction. The yearly unemployment rate is measured at the regional level where a detainee is released. Over our sample period, it varies between 4.4 and 17.5%.

As stressed above, a host of legislative and policy changes designed to "tackle crime" was implemented in 2012. In order to identify the impact of the new policy, we distinguish between two groups: Those who entered the panel before the implementation and those who entered after. We refer to the former as `Early` entrants

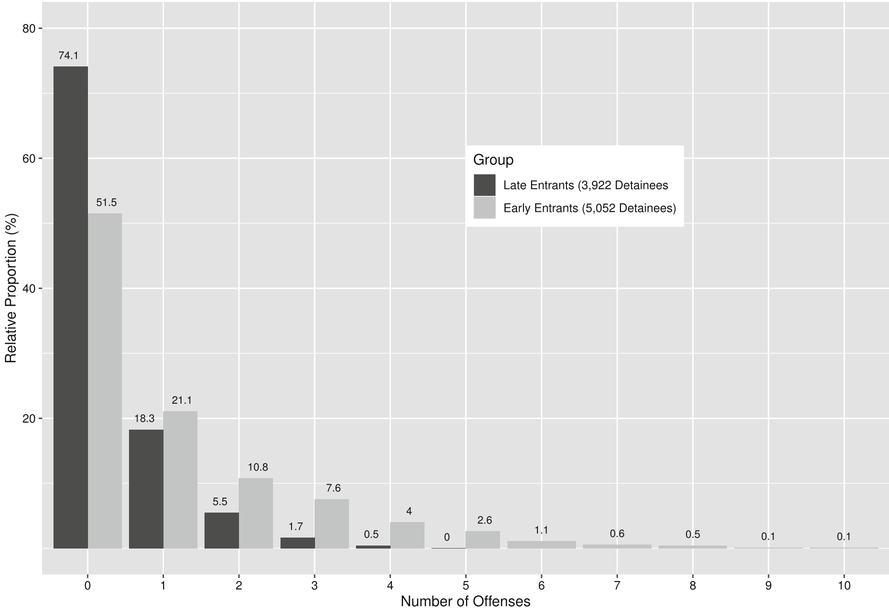

Fig. 2 Frequency of repeat offenses

(5052 detainees) and to the latter as `Late` entrants (3922 detainees). `Early` entrants were thus exposed to two policy regimes (before and after 2012), while `Late` entrants were only exposed to the tough-on-crime regime. Because the `Early` entrants are observed for a longer period, there are more observations in the post-2012 period than before. As reported in the table, roughly 75% of the sample observations were exposed to the tough-on-crime regime. Thus the `Early` group was exposed to both the pre- and post-2012 periods, whereas the `Late` group was only exposed to the tough-on-crime period. The next 3 lines of the table provide information on the yearly rates of recidivism for distinct periods.[10] Thus, the overall rate of recidivism is equal to 11.4%. `Early` entrants have a recidivism rate of 9%, while `Late` entrants have a much lower rate at 2.3%, partly because they are observed for a shorter period of time.

Figure 2 depicts the proportions of repeat offenses for the entire sample period and for those who entered the panel in 2012 or later. The figure provides *prima facie* evidence on the impact of the policy. Indeed, the proportion of detainees who do not reoffend upon release in the post-policy period is 15 percentage points larger (74.1%) than the proportion for the whole sample period (51.5%). Likewise, the proportion of repeat offenders is between 3 to 6 percentage points lower in the post-policy period for any given number of repeat offenses.[11] Naturally, such differences may results

[10] Recidivism is a yearly dummy variable equal to one the year at which the new incarceration begins and zero otherwise. Recidivism may be equal to one in consecutive years so long as the repeat offenses occurred after the end of the previous sentence. Reincarcerations while on parole or on conditional release are not considered repeat offenses.

[11] Obviously, detainees who entered the sample on or after 2012 have had less time to reoffend. Yet, in our sample as many as 34% of detainees are reincarcerated within 12 months upon release, and as many as

from factors other than the "tough-on-crime" policy, such as, but not limited to, better economic opportunities, and demographic compositional changes. In order to net these out, we now turn to formal econometric modeling.[12]

5.2 Estimation results

The dependent variable y is a binary variable that equals 1 if an individual commits a repeat offense and 0 otherwise. Since the dependent variable is binary (discrete, more generally), it cannot yield continuous quantiles. Our concern is to model the latent utility differential between committing and not committing a repeat offense, within a quantile framework. This may be interpreted as some sort of propensity or willingness to commit a repeat offense. The list of covariates can be classified into time-varying covariates (age, schooling, unemployment rate), time-invariant policy variables (`Early-Pre2012`, `Early-Post2012` and `Late`) and other time-invariant control variables. Since `Early` entrants are observed both before and after 2012, we can use two period-specific effects to investigate their responsiveness to the implementation of the policy. Likewise, since we do not include an intercept in the model, we can separately estimate the relative sensitivity of the `Late` group.

Our Bayesian setup uses the same independent prior distributions as in the simulation exercise: $\beta \sim N_k\left(0_k, 10^3 I_k\right)$, $\zeta \sim N_3\left(0_3, 10^3 I_3\right)$, $\sigma_\alpha^2 \sim IG\,(10/2, 9/2)$. We generate 60, 000 MCMC samples of which the first 10, 000 are discarded as burn-ins. The posterior estimates are reported using a thinning factor of 50, optimized following the approach in Owen (2017).[13]

The mixing of the MCMC chain is extremely good as illustrated in Fig. 3 which exhibits the trace plots of the parameters at the 75th quantile.[14] Trace plots at other quantiles are similar and not reported for the sake of brevity, but they are available upon request. Figure 4 provides additional information on the performance of the MCMC chain. The left-hand-side figure depicts the boxplots of the inefficiency factors of the parameters (βs, ζs and σ_α^2) for each of the five different quantiles used in estimating the model. Except perhaps for the 10th quantile, all are reasonably close to one. Consistent with the simulation results, the parameter with the largest inefficiency factor at the 10th quantile is σ_α^2 (not shown, see Table 2). The right-hand-side figure reports the boxplots of the convergence diagnostics of the parameter estimates for the same five specifications based on the first 10% and the last 40% values of the Markov

Footnote 11 continued
43% within 2 years. Hence, the sharp decline in repeat offenses in the post-2012 period is unlikely due to the sampling frame. See Lalande et al. (2015).

[12] To the extent the new legislation has indeed lowered the recidivism rates, it not clear whether it did so through deterrent or incapacitative effects. Yet, see Bhuller et al. (2020) for US evidence according to which deterrence dominates incapacitation.

[13] Thinning has been criticized by some (MacEachern and Berliner 1994; Link and Eaton 2012), while others acknowledge that it can increase statistical efficiency (Geyer 1991). See Owen (2017) who claims that the arguments against thinning may be misleading.

[14] Note that the time-varying covariates (`Age`, `Schooling` and `Unemployment rate`) have been "demeaned" and that `Age` has been divided by 10. The parameter estimates must thus be interpreted accordingly.

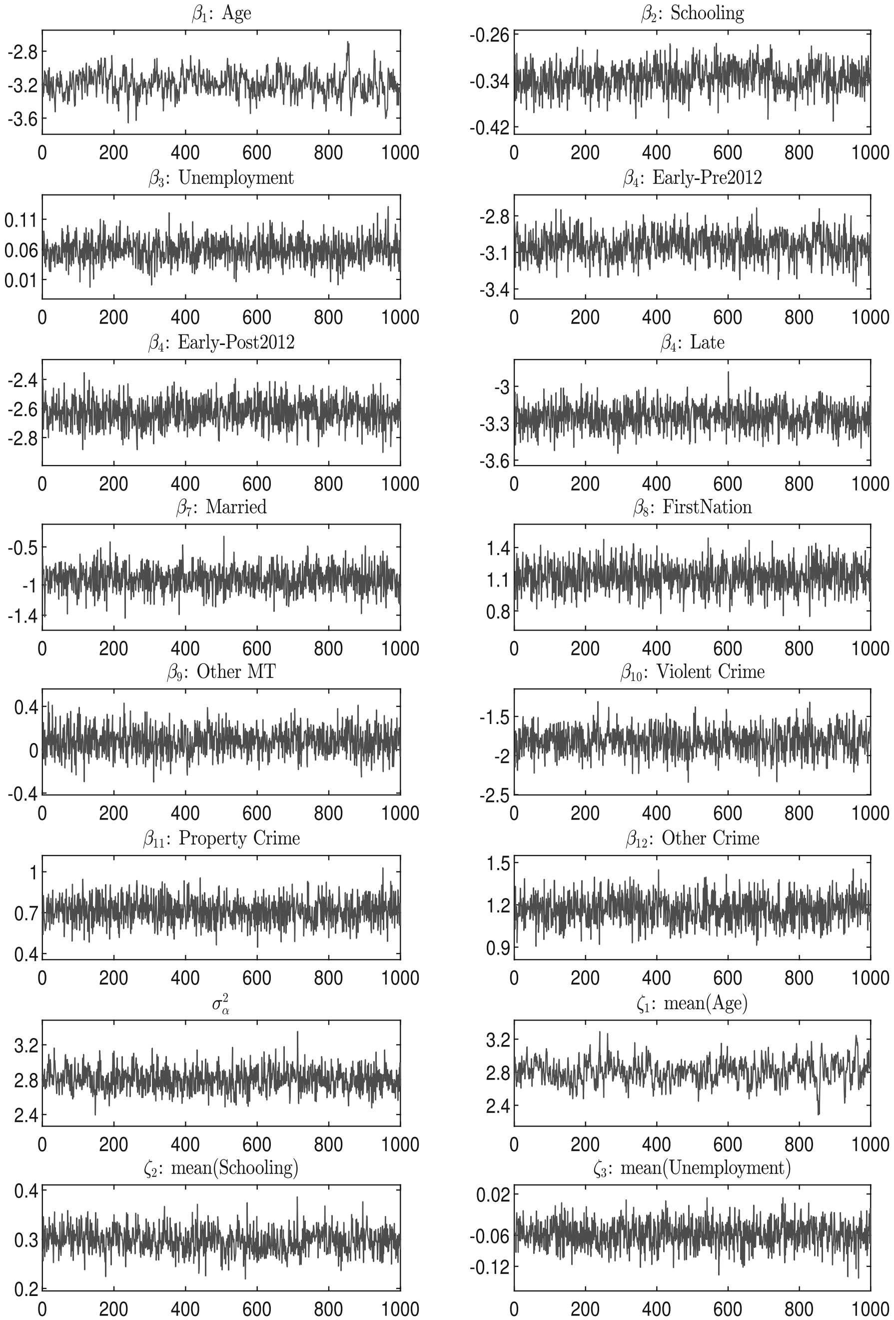

Fig. 3 Trace plots of the parameters for the 75th quantile

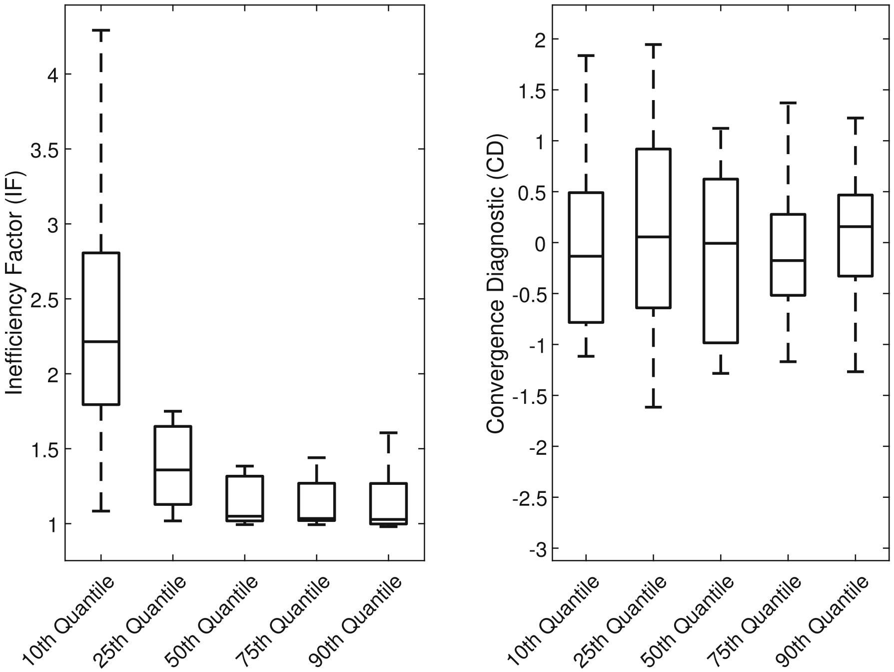

Fig. 4 Boxplots of the inefficiency factor and convergence diagnostic for $(\beta, \zeta, \sigma_\alpha^2)$ of the parameters from quantile regression at the 10th, 25th, 50th, 75th, and 90th quantiles

chain (Geweke 1992). As depicted, all parameters have Z-scores within 1.96 standard deviation of the mean at the 5% level or within 2.58 standard deviation at 1% level. All in all, the Markov chains behave satisfactorily and thus lend themselves to statistical inference.

Table 5 reports the posterior means and standard deviations of the parameters at five different quantiles separately. To ease interpretation, the quantile-specific estimates are reported column-wise in increasing order. Row-wise, we distinguish the time-varying covariates from the time-invariant and the CRE variables. Note that the CRE specification does not include an intercept. As stated above, this is to allow the identification of the three time-invariant policy variables, `Early-Pre2012`, `Early-Post2012` and `Late`. The two former are equal to one if the detainee entered the panel before 2012 but was incarcerated prior or after 2012, respectively. The latter is equal to one if a detainee's first incarceration occurred during or after 2012, and thus always exposed to the tough-on-crime policy. All other time-invariant variables are measured at first entry in the panel.[15] The estimates of the correlated random components associated with the individual mean `Age`, `Schooling` and `Unemployment`, $\widehat{\zeta}$, are all statistically different from zero regardless of the quantile. The individual-specific effects, α_i, are thus highly correlated with the individual means of the time-varying variables.

[15] Recall from Table 4 that very few men are married. In addition, next to none report a change in their marital status in between incarcerations. Further, since the marital status of nonrepeaters is not observed in the data, we are constrained to use the information at entry in the panel.

Table 5 Posterior mean (Mean) and standard deviation (SD) of the parameters in the crime recidivism model

Variable	10th quantile		25th quantile		50th quantile		75th quantile		90th quantile	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Time varying covariates										
β-age	−14.797	0.618	−6.780	0.274	−3.805	0.149	−3.191	0.136	−4.528	0.197
β-schooling	−1.464	0.114	−0.647	0.047	−0.380	0.025	−0.336	0.023	−0.479	0.031
β-unemp rate	0.364	0.105	0.152	0.042	0.077	0.023	0.059	0.019	0.082	0.028
Policy variables (Time invariant)										
Early-pre2012	−29.222	0.525	−11.966	0.239	−5.446	0.125	−3.044	0.106	−2.171	0.149
Early-post2012	−27.593	0.450	−11.009	0.207	−4.902	0.109	−2.627	0.090	−1.668	0.123
Late	−30.906	0.515	−12.461	0.227	−5.680	0.121	−3.245	0.099	−2.556	0.135
Other time invariant covariates										
Married	−5.321	0.920	−2.311	0.405	−1.224	0.205	−0.925	0.158	−1.292	0.219
Aboriginals	5.586	0.665	2.494	0.271	1.361	0.146	1.127	0.130	1.650	0.190
Oth. Mot. Ton.	0.311	0.692	0.183	0.283	0.100	0.151	0.078	0.119	0.095	0.165
Violent crime	−12.889	0.875	−4.946	0.408	−2.481	0.214	−1.816	0.161	−2.262	0.209
Property crime	2.561	0.459	1.636	0.212	0.914	0.112	0.713	0.091	0.972	0.123
Other crime	5.095	0.478	2.676	0.219	1.468	0.112	1.169	0.095	1.642	0.128
Correlated random effects										
ζ-age	12.530	0.629	5.879	0.281	3.336	0.152	2.809	0.137	3.986	0.198
ζ-schooling	1.251	0.126	0.559	0.051	0.333	0.027	0.298	0.024	0.422	0.033
ζ-unemployment	−0.331	0.141	−0.142	0.057	−0.073	0.030	−0.058	0.025	−0.088	0.036
σ_α^2	75.610	3.165	13.218	0.537	3.909	0.165	2.803	0.129	6.268	0.329

Omitting this correlation may therefore bias the model estimates and hence their intrinsic marginal effects and relative risks. This provides empirical support to the inclusion of CRE within a quantile regression model.

The first noteworthy feature of the table is that all parameter estimates are statistically different from zero, except for the parameter associated with `Other Mother Tongue`. Thus detainees who report speaking a language other than English or French at home are no more and no less likely to eventually reoffend. A second interesting feature concerns the sign of the parameter estimates. Indeed, all are consistent with recent research on crime recidivism. For instance, `Age` and `Schooling` are associated with lower rates of recidivism (Bhuller et al. 2020), whereas being released during a period of high unemployment has been found to favor recidivism (Siwach 2018; Rege et al. 2019). Likewise, married men are less likely to reoffend, whereas Aboriginal detainees are more likely to do so (Justice Canada 2017). The type of crime is also associated with recidivism. The estimates must be interpreted relative to traffic-related crimes, which is the base or omitted category in our analysis. Clearly, sentences for `Violent Crimes` will be harsher and so the large parameter estimate presumably reflects an incapacitation effect. Finally, the policy parameters `Early-Pre2012`, `Early-Post2012`, and `Late` are all negative and statistically different from zero. Note that the estimates of `Early-Post2012` are all slightly smaller than those of `Early-Pre2012` as well as those of `Late`.

As stated in Sect. 4, the parameter estimates such as those reported in Table 5 do not give the marginal effects. In addition, it does not follow that the magnitudes of the parameter estimates are a good proxy for their marginal effects. Yet, the latter are important from a policy perspective. Thus, while the parameter estimates vary considerably across quantiles, it is not clear that the marginal effects are equally sensitive since they depend both on the time-varying variables and the correlated random components. Figure 5 reports the average marginal effects computed according to Eq. (10), along with their highest posterior density intervals (HPDI).[16] Note that most marginal effects have a relatively flat profile between the 10th and 75th quantiles and then exhibit a small kink between the 75th and 90th quantiles. For instance, increasing `Age` by 1/10th reduces the probability of reoffending by 1% at the 10th quantile and by 2.0% at the 90th quantile. Similar results hold for `Schooling` (1% *vs* 2.0%), `Married` (0.3% *vs* 0.55%), and `Violent Crime` (6% *vs* 7.7%). Thus, for all three time-varying covariates the marginal effects increase by one half as we move from the 10th to the 90th quantile. As for the time-invariant variables, their marginal effects all increase by at least 50% as we move from the 10th to the 90th quantile. In particular, the marginal effects associated to `First Nation`, `Property Crime`, and `Other Crime` exhibit a twofold increase. More importantly, the marginal effects of the three "tough-on-crime" variables increase manifold and in a steady fashion between the 10th and 90th quantiles. Furthermore, the HPDI is relatively narrow in all three cases. Hence, according to the parameter estimates associated with `Early-Pre2012` the probability of reoffending decreases from 28% at the 10th quantile to as little as 5%

[16] The marginal effects for `Age` correspond to 1/10 of an additional year relative to the mean. Those for `Unemployment` and `Schooling` correspond to one additional year and one additional percentage point relative to their individual means, respectively. The remaining marginal effects correspond to a change in the indicator variables.

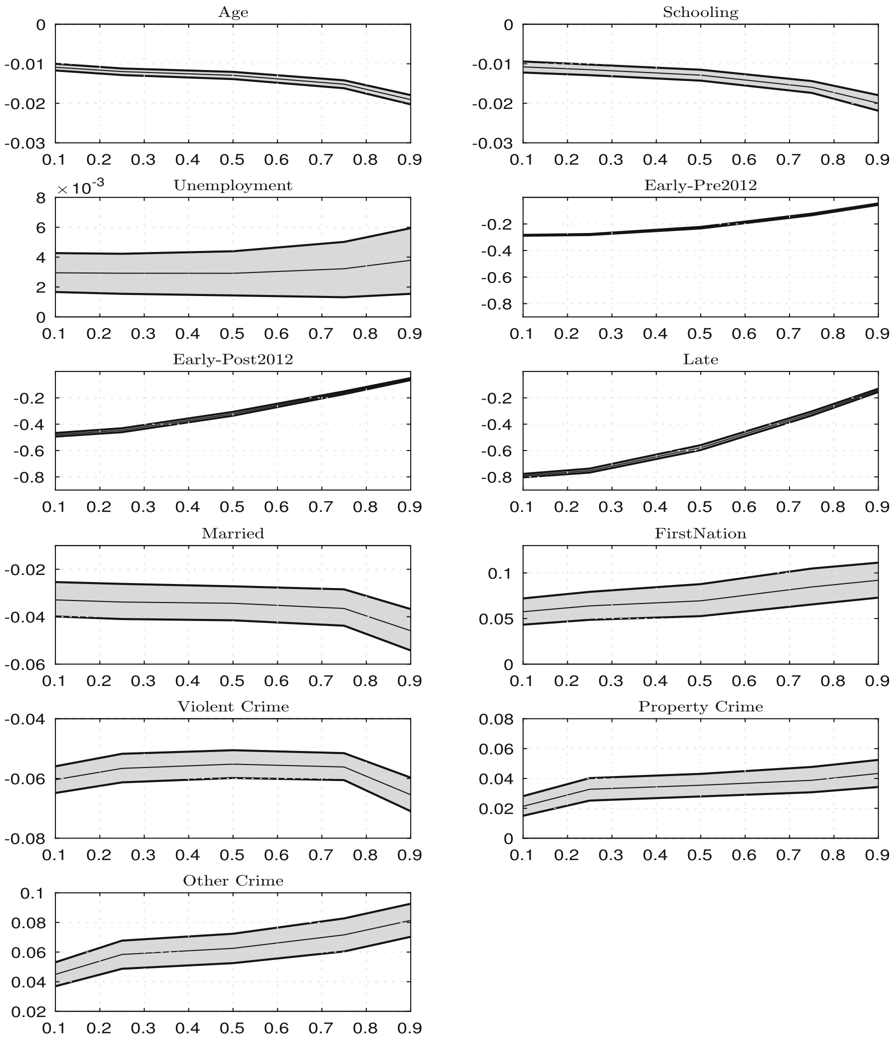

Fig. 5 Marginal effects with 95% HPDI

at the 90th. Interestingly, the probability of reoffending of the `Early-Post2012` decreases just as much, from 48% at the 10th quantile to as little as 5% at the 90th. Likewise, the parameters of `Late` group imply that the probability decreases from 79 to 14% at both extremes.

To highlight these results, we report the marginal effects of the policy variables in Fig. 6 using the same scale. As is readily shown, all three profiles are upward sloping and have distinct HPDIs, except for the Early-Pre and Early-Post groups at the 90th quantile. Note that Early entrants are more inclined to reoffend in the pre-2012 period than in the post-2012 period. This result is consistent with the literature that finds that imprisonment discourages further criminal behavior, and that the reduction extends

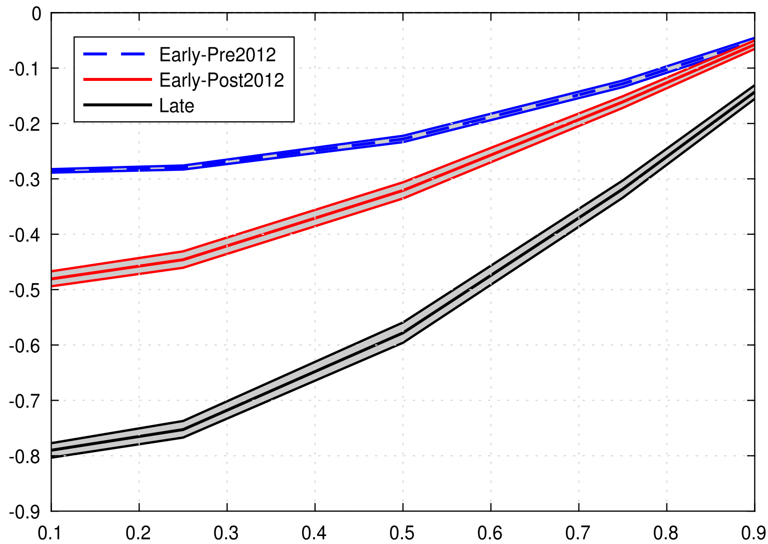

Fig. 6 Marginal effects of policy variables with 95% HPDI

beyond incapacitation (Bhuller et al. 2020). On the other hand, this result does not run counter to the literature on peer effects according to which prison interactions are conducive to further recidivism (Bayer et al. 2009; Stevenson 2017; Marchand 2020). Indeed, absent the "tough-on-crime" policy, the marginal effects profile of the `Early-Post` group may have lied above that of the `Early-Pre` group. The fact that it lies below suggests either that the disincentive effect dominates in the longer run (e.g. Bhuller et al. 2020) or that the "tough-on-crime" policy dominates potential peer effects. All in all, these results are important from a policy perspective for three reasons. First, they imply that detainees are sensitive to the "tough-on-crime" policy. Second, since `Early` entrants are less likely to reoffend in the post-2012 period, it appears that the "tough-on-crime" policy (and/or past incarceration) has had a significant impact on recidivism. This is also evidenced by large differences between the `Early-Pre2012` and `Late` profiles. Third, and most importantly, the policy does not impact all detainees alike. Those in the lower quantiles are much more responsive than those in the upper quantiles.

In order to gain further insight into the sensitivity of recidivism to various covariates, we report the corresponding relative risks in Fig. 7 (see Eq. (11)) along with their HDPI. Not surprisingly given the marginal effects, the relative risks are fairly constant for the first two or three quantiles ($p = 10\%, 25\%, 50\%$), with a few exceptions. Beyond the second or third quantiles, most increase or decrease sharply. The figure also shows which covariates influence recidivism most. Thus, while `Age`, `Schooling`, and `Unemployment Rate` are associated with slightly different rates of repeat offenses, only those in the highest quantiles exhibit statistically different recidivism rates. On the other hand, marital status (`Married`), `First Nation`, and types of crime (`Violent, Property, Other`) all have statistically higher or lower relative risks of reoffending as the case may be, and all exhibit a sharp change between

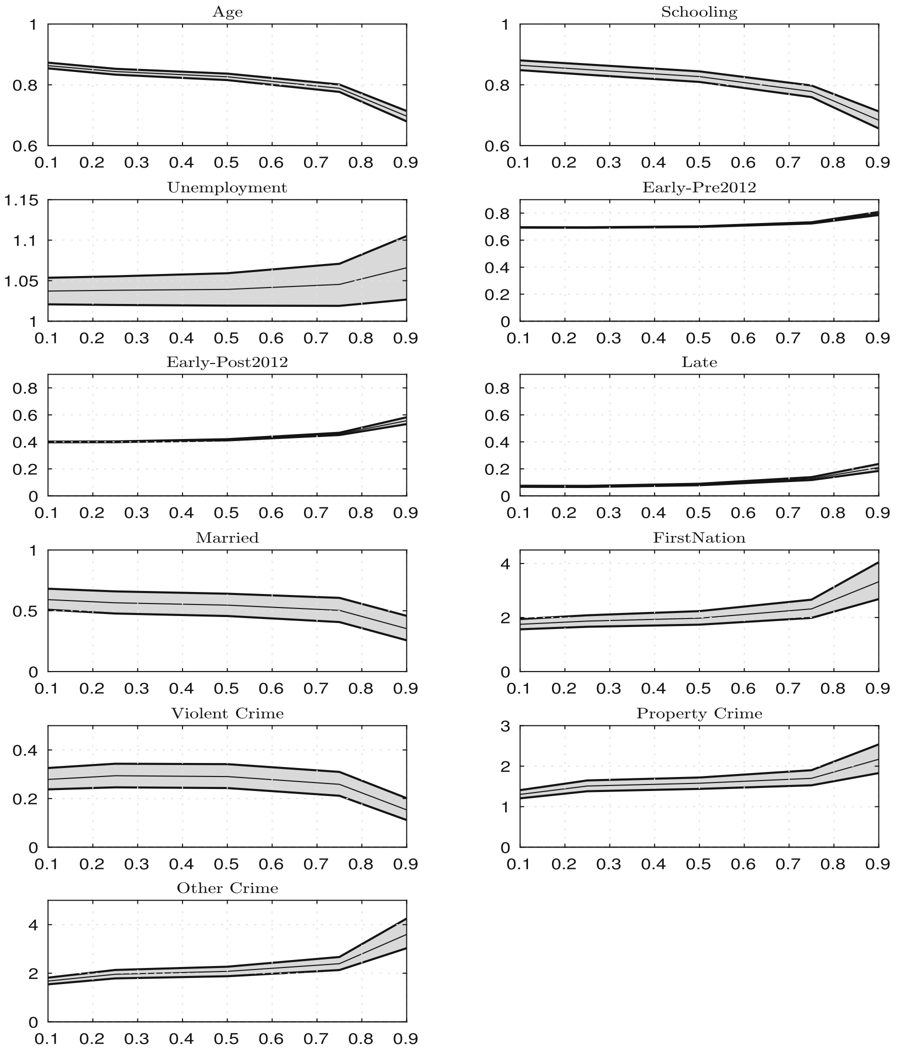

Fig. 7 Relative risks with 95% HPDI

the last two quantiles. Here, as with the previous figure, the results concerning the "tough-on-crime" variables are particularly interesting. Indeed, according to the figure all detainees were much less likely to reoffend in the post 2012 period, irrespective of whether they where first convicted prior to 2012 or after. As with the marginal effects, the policy appears to have had a larger impact on those in the lower quantiles.

To better ascertain the impact of the policy variables, we report the relative risks of the three groups in Fig. 8. The two groups who were exposed to the tough-on-crime policy (`Late` and `Early-Post`) have the lowest profiles. For instance, the 95% HPDI at the 10th quantile is [0.692; 0.696] and [0.396; 0.402] for the `Early-Pre` and `Early-Post` groups, respectively, and [0.066; 0.075] for the `Late group`. On

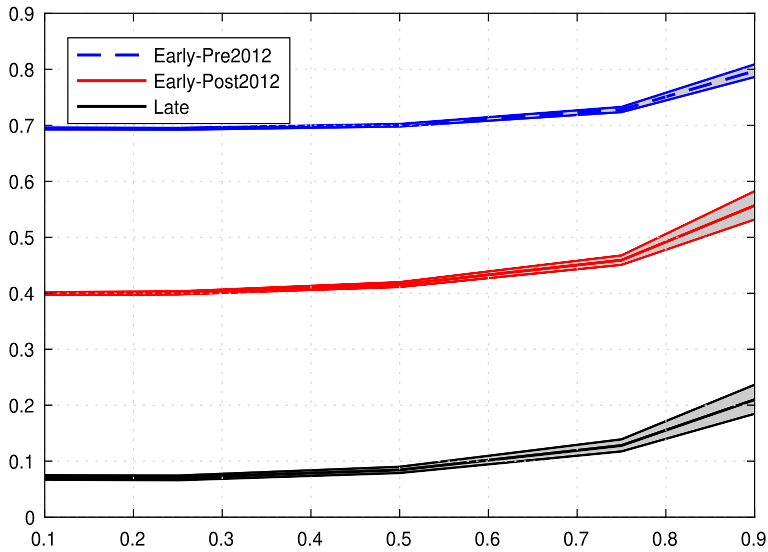

Fig. 8 Relative risks of policy variables with 95% HPDI

the other hand, the 95% HPDI at the 90th quantile for the three groups is [0.786; 0.808], [0.531; 0.582], and [0.184; 0.236] , respectively. In other words, for those in the lowest quantile, exposure to the policy decreases recidivism by as much as [30.4; 30.8] %, [59.8; 60.4] %, and [92.5; 93.4] %, respectively. In contrast, for those in the highest quantile, the decrease in the relative risk is still sizeable, albeit somewhat smaller. But the main lesson of this exercise is the highly significant effect of the "tough-on-crime" policy on the probability of recidivism. The figures also highlight the important but sometimes omitted fact that the response to a given public policy can be quite heterogeneous.

6 Conclusion

This paper presents a panel quantile regression model for binary outcomes with correlated random effects (CRE) and proposes two MCMC algorithms for its estimation. By incorporating the CRE into the panel quantile regression for discrete outcomes, we move beyond the random effects framework typically considered in the Bayesian quantile regression literature. The paper makes an important contribution to the literature on quantile regression for panel data and panel quantile regression for discrete outcomes. The two proposed MCMC algorithms are simpler to implement, but we prefer the algorithm that exploits block sampling of parameters to reduce the autocorrelation in MCMC draws. This blocked algorithm is tested in multiple simulation studies and shown to perform extremely well. We also emphasize the calculation of marginal effects in models with discrete outcome and explain its computation, along with those of relative risk and odds ratio, using the MCMC draws. Finally, we implement the proposed quantile framework to analyze crime recidivism in Quebec (a Canadian Province) for the period 2007–2017 using novel data from administrative correctional files. Among other things, we investigate the effect of the recently implemented “tough-on-crime” policy on the probability of repeat offense. Our results show that the policy negatively affects the probability of repeat offenses across quantiles and hence has been largely successful in achieving its objective. Besides, the results suggest that the CRE structure is relevant in modeling the probability of repeat offenses across quantiles.

This paper opens avenues for future research in several directions. The proposed framework can be readily extended to panel quantile regression models with continuous and other discrete response variables (e.g., count and ordinal outcomes). One may also consider the Hausman–Taylor version of CRE, where the individual-specific effects are related to only some of the time-varying and time-invariant regressors, and merge it with the panel quantile regression model for continuous or discrete outcomes. Besides, a dynamic relationship can be introduced to panel quantile regression models (with continuous or discrete outcomes) and the initial condition problems can be tackled using the CRE structure.

Acknowledgements This paper is written in honor of Professor Badi H. Baltagi for his valuable contributions to econometrics. We are grateful to Bernard Chéné, Senior Advisor, Programs Directorate, Public Safety (Québec), for his advice and for granting us access to the data used in the paper. We are also grateful to William Arbour, Steeve Marchand, and Ivan Jeliazkov for their advice and numerous discussions. Finally, we are indebted to the editors Qi Li, Vasilis Sarafidis, and Joakim Westerlund and to two anonymous referees for their useful comments and suggestions which helped us in substantially improving the manuscript. The usual disclaimers apply.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abrevaya J, Dahl CM (2008) The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. J Bus Econ Stat 26(4):379–397

Albert J, Chib S (1993) Bayesian analysis of binary and polychotomous response data. J Am Stat Assoc 88(422):669–679

Alhamzawi R (2016) Bayesian model selection in ordinal quantile regression. Comput Stat Data Anal 103:68–78

Alhamzawi R, Ali HTM (2018) Bayesian quantile regression for ordinal longitudinal data. J Appl Stat 45(5):815–828

Alhamzawi R, Ali HTM (2020) Bayesian single-index quantile regression for ordinal data. Commun Stat Simul Comput 49(5):1306–1320

Arellano M (1993) On the testing of correlated effects with panel data. J Econom 59(1–2):87–97

Arellano M, Bonhomme S (2016) Nonlinear panel data estimation via quantile regression. Econom J 19(3):61–94

Bache SHM, Dahl CM, Christensen JT (2013) Headlights on tobacco road to low birthweight outcomes: evidence from a battery of quantile regression estimators and a heterogeneous panel. Empir Econ 44(3):1593–1633

Baltagi BH (2006) Estimating an economic model of crime using panel data from North Carolina. J Appl Econom 21(4):543–547

Baltagi BH (2013) Econometric analysis of panel data, 5th edn. Wiley, Chichester

Baltagi BH, Bresson G, Pirotte A (2003) Fixed effects, random effects or Hausman–Taylor?: A pretest estimator. Econom Lett 79(3):361–369
Baltagi BH, Bresson G, Chaturvedi A, Lacroix G (2018) Robust linear static panel data models using ϵ-contamination. J Econom 202(1):108–123
Barrodale I, Roberts FDK (1973) Improved algorithm for discrete l_1 linear approximation. SIAM J Numer Anal 10(5):839–848
Bayer P, Hjalmarsson R, Pozen D (2009) Building criminal capital behind bars: peer effects in juvenile corrections. Q J Econ 124(1):105–147
Benoit DF, Poel DVD (2010) Binary quantile regression: a Bayesian approach based on the asymmetric Laplace distribution. J Appl Econom 27(7):1174–1188
Bhuller M, Dahl G, Loken K, Mogstad M (2020) Incarceration, recidivism and employment. J Polit Econ 128:1269–1324
Burda M, Harding M (2013) Panel probit with flexible correlated effects: quantifying technology spillovers in the presence of latent heterogeneity. J Appl Econ 28(6):956–981
Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge University Press, Cambridge
Canay IA (2011) A simple approach to quantile regression for panel data. Econom J 14(3):368–386
Chalfin A, McCrary J (2017) Criminal deterrence: a review of the literature. J Econ Lit 55(1):5–48
Chamberlain G (1980) Analysis with qualitative data. Rev Econ Stud 47:225–238
Chamberlain G (1982) Multivariate regression models for panel data. J Econom 18(1):5–46
Chamberlain G (1984) Panel data. In: Griliches Z, Intriligator MD (eds) Handbook of econometrics, vol 2. Elsevier, Amsterdam, pp 1247–1318
Chen C (2007) A finite smoothing algorithm for quantile regression. J Comput Graph Stat 16(1):136–164
Chernozhukov V, Fernández-Val I, Hahn J, Newey W (2013) Average and quantile effects in nonseparable panel models. Econometrica 81(2):535–580
Chib S, Carlin BP (1999) On MCMC sampling in hierarchical longitudinal models. Stat Comput 9:17–26
Chib S, Jeliazkov I (2006) Inference in semiparametric dynamic models for binary longitudinal data. J Am Stat Assoc 101(474):685–700
Cornwell C, Trumbull WN (1994) Estimating the economic model of crime with panel data. Rev Econ Stat 76(2):360–366
Dantzig GB (1963) Linear programming and extensions. Princeton University Press, Princeton
Dantzig GB, Thapa MN (1997) Linear programming 1: introduction. Springer, New York
Dantzig GB, Thapa MN (2003) Linear programming 2: theory and extensions. Springer, New York
Davino C, Furno M, Vistocco D (2013) Quantile regression: theory and applications. Wiley, Chichester
Davis CS (1991) Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. Stat Med 10(12):1959–1980
Devroye L (2014) Random variate generation for the generalized inverse Gaussian distribution. Stat Comput 24(2):239–246
Galvao AF, Kato K (2017) Quantile regression methods for longitudinal data. In: Koenker R, Chernozhukov V, He X, Peng L (eds) Handbook of quantile regression. Chapman and HAll/CRC, New York, pp 363–380
Galvao AF, Poirier A (2019) Quantile regression random effects. Ann Econ Stat 134:109–148
Galvao AF, Lamarche C, Lima LR (2013) Estimation of censored quantile regression for panel data with fixed effects. J Am Stat Assoc 108(503):1075–1089
Geraci M, Bottai M (2007) Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics 8(1):140–154
Geraci M, Bottai M (2014) Linear quantile mixed models. Stat Comput 24:461–479
Geweke J (1991) Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities, Iowa City, IA, USA. http://www.biz.uiowa.edu/faculty/jgeweke/papers/paper47/paper47.pdf
Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian statistics, vol 4. Clarendon Press, Oxford, pp 169–193
Geweke J (2005) Contemporary Bayesian econometrics and statistics. Wiley, Chichester
Geyer CJ (1991) Markov chain Monte Carlo maximum likelihood. In: Kemramides EM (ed) Computing science and statistics: proceedings of the 23rd symposium on the interface. Interface Foundation of North America, Fairfax Station, VA, USA, pp 156–163

Ghasemzadeh S, Ganjali M, Baghfalaki T (2018) Bayesian quantile regression for analyzing ordinal longitudinal responses in the presence of non-ignorable missingness. METRON 76(3):321–348
Ghasemzadeh S, Ganjali M, Baghfalaki T (2020) Bayesian quantile regression for joint modeling of longitudinal mixed ordinal and continuous data. Commun Stat Simul Comput 49(2):375–395
Gibbons RD, Hedeker D (1993) Application of random effects probit regression. J Consult Clin Psychol 62(2):285–296
Graham BS, Hahn J, Poirier A, Powell JL (2018) A quantile correlated random coefficients panel data model. J Econom 206(2):305–335
Greenberg E (2012) Introduction to Bayesian econometrics, 2nd edn. Cambridge University Press, New York
Greene W (2015) Panel data models for discrete choice. In: Baltagi BH (ed) The Oxford handbook of panel data. Oxford University Press, New York
Greene WH (2017) Econometric analysis, 8th edn. Prentice Hall, New York
Gu J, Volgushev S (2019) Panel data quantile regression with grouped fixed effects. J Econom 213(1):68–91
Harding M, Lamarche C (2017) Penalized quantile regression for semiparametric models with correlated individual effects. J Appl Econom 32(2):342–358
Hausman JA (1978) Specification tests in econometrics. Econometrica 46(6):1251–1271
Hausman JA, Taylor WE (1981) Panel data and unobservable individual effects. Econometrica 49(6):1377–1398
Jeliazkov I, Rahman MA (2012) Binary and ordinal data analysis in economics: modeling and estimation. In: Yang XS (ed) Mathematical modeling with multidisciplinary applications. Wiley, New York, pp 123–150
Jeliazkov I, Vossmeyer A (2018) The impact of estimation uncertainty on covariate effects in nonlinear models. Stat Pap 59(3):1031–1042
Jeliazkov I, Graves J, Kutzbach M (2008) Fitting and comparison of models for multivariate ordinal outcomes. Adv Econom Bayesian Econom 23:115–156
Joshi R, Wooldridge JM (2019) Correlated random effects models with endogenous explanatory variables and unbalanced panels. Ann Econ Stat 134:243–268
Justice Canada (2017) Indigenous overrepresentation in the criminal justice system. https://www.justice.gc.ca/eng/rp-pr/jr/jf-pf/2017/docs/jan02.pdf
Karmarkar N (1984) A new polynomial time algorithm for linear programming. Combinatorica 4(4):373–395
Kobayashi G, Kozumi H (2012) Bayesian analysis of quantile regression for censored dynamic panel data model. Comput Stat 27(2):359–380
Koenker R (2004) Quantile regression for longitudinal data. J Multivar Anal 91(1):74–89
Koenker R (2005) Quantile regression. Cambridge University Press, Cambridge
Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46(1):33–50
Koenker R, d'Orey V (1987) Computing regression quantiles. J R Stat Soc Ser C 36(3):383–393
Kordas G (2006) Smoothed binary regression quantiles. J Appl Econom 21(3):387–407
Kozumi H, Kobayashi G (2011) Gibbs sampling methods for Bayesian quantile regression. J Stat Comput Simul 81(11):1565–1578
Lalande P, Pelletier Y, Dolmaire P, Raza E (2015) Projet, enquête sur la récidive/reprise de la clientèle confiée aux services correctionnels du Québec. Ministère de la sécurité publique du Québec (http://collections.banq.qc.ca/ark:/52327/2505967)
Lamarche C (2010) Robust penalized quantile regression estimation for panel data. J Econom 157(2):396–408
Link WA, Eaton MJ (2012) On thinning of chains in MCMC. Methods Ecol Evol 3:112–115
Liu Y, Bottai M (2009) Mixed-effects models for conditional quantiles with longitudinal data. Int J Biostat 5(1):1–24
Luo Y, Lian H, Tian M (2012) Bayesian quantile regression for longitudinal data models. J Stat Comput Simul 82(11):1635–1649
MacEachern SN, Berliner LM (1994) Subsampling the Gibbs sampler. Am Stat 48(3):188–190
Madsen K, Nielsen HB (1993) A finite smoothing algorithm for linear l_1 estimation. SIAM J Optim 3(2):223–235
Marchand S (2020) Peer effects in prison and recidivism. Mimeo, University of California, Berkeley
Mehrotra S (1992) On the implementation of primal-dual interior point methods. SIAM J Optim 2(4):575–601

Mundlak Y (1978) On the pooling of time series and cross section data. Econometrica 46(1):69–85
Omata Y, Katayama H, Arimura TH (2017) Same concerns, same responses: a Bayesian quantile regression analysis of the determinants for nuclear power generation in Japan. Environ Econ Policy Stud 19(3):581–608
Owen AB (2017) Statistically efficient thinning of a Markov chain sampler. J Comput Graph Stat 26(3):738–744
Rahman MA (2013) Quantile regression using metaheuristic algorithms. Int J Comput Econ Econom 3(3/4):205–233
Rahman MA (2016) Bayesian quantile regression for ordinal models. Bayesian Anal 11(1):1–24
Rahman MA, Karnawat S (2019) Flexible Bayesian quantile regression in ordinal models. Adv Econom 40B:211–251
Rahman MA, Vossmeyer A (2019) Estimation and applications of quantile regression for binary longitudinal data. Adv Econom 40(B):157–191
Rege M, Skardhamar T, Telle K, Votruba M (2019) Job displacement and crime: evidence from Norwegian register data. Labour Econ 61:101761
Siwach G (2018) Unemployment shocks for individuals on the margin: exploring recidivism effects. Labour Econ 52:231–244
Soares YM, Fagundes RA (2018) Interval quantile regression models based on swarm intelligence. Appl Soft Comput 72:474–485
Stevenson M (2017) Breaking bad: mechanisms of social influence and the path to criminality in juvenile jails. Rev Econ Stat 99(5):824–838
Wang J (2012) Bayesian quantile regression for parametric nonlinear mixed effects models. Stat Methods Appl 21(3):279–295
Wooldridge JM (2010) Econometric analysis of cross section and panel data, 2nd edn. MIT Press, Cambridge
Yu K, Moyeed RA (2001) Bayesian quantile regression. Stat Probab Lett 54(4):437–447
Yuan Y, Yin G (2010) Bayesian quantile regression for longitudinal studies with nonignorable missing data. Biometrics 66(1):105–114