Latent class multi-state quantile regression with a cure fraction: application to jail recidivism in the U.S.

Rosario Barone, Alessio Farcomeni

Journal of the Royal Statistical Society Series A: Statistics in Society, 2026, **189**, 1697–1717
https://doi.org/10.1093/jrsssa/qnaf139
Advance access publication 15 September 2025
Original Article

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

Latent class multi-state quantile regression with a cure fraction: application to jail recidivism in the U.S.

Rosario Barone[1] and **Alessio Farcomeni**[2]

[1]Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Largo Gemelli 1, Milan 20123, Italy
[2]Department of Economics and Finance, University of Rome Tor Vergata, Via Columbia 2, Rome 00133, Italy

Address for correspondence: Rosario Barone, Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Largo Gemelli 1, Milan 20123, Italy. Email: rosario.barone@unicatt.it

Abstract

We propose a multi-state quantile regression model that admits a cure-fraction for each possible transition, so that individuals may not experience that event. A discrete latent variable allows us to take into account unobserved heterogeneity. The model is estimated in a Bayesian framework, without specifying the number of latent classes. A simple strategy to scale inference to big data is discussed. We are motivated by an original application to jail recidivism in the U.S. between 2020 and 2023. We find that 20% of the subjects have high cumulative hazard of recidivism; with little association to covariates such as age, gender, crime, and ethnicity. A latent group has been shown to accumulate up to two detentions per year of freedom and represents about 10% of the population.

Keywords: cumulative hazard, Dirichlet process prior, unobserved heterogeneity

1 Introduction

Research on incarceration has traditionally focused on prisons, but the scale of admissions to jails is significantly much higher. In the U.S., prisons are generally under the jurisdiction of the state or federal government, where convicted offenders often serve sentences longer than 1 year. Jails instead are local facilities where individuals are held while awaiting trial, or serving shorter sentences. In our data, observed jail detentions are very short (an application of our model indicates that the median duration for jail detention is 3 weeks, 75% of jail detentions last less than 2 months, and 90% less than six). Even if most jail detentions are short, there is a very high lifetime risk of recidivism in general, and at least one detention will be experienced by a surprisingly high proportion of the U.S. population (Western et al., 2021). Jail detention spells have been found to adversely affect court outcomes, earnings, and family life (Western et al., 2021). It is also true that earnings affect detention, which is then an issue of economic inequality: according to Sawyer and Wagner (2022), 67% of individuals are in jail awaiting trial because they have insufficient wealth to support the payment of a bail.

The correctional role of jails is unclear. Recidivism, defined as a new period of detention after release, is a key end point for evaluating the effectiveness of social policies, programs, and initiatives (e.g. employment and/or education actions), and the deterrent effect of the sentences themselves (Beaudry et al., 2021; Golder et al., 2005; Lange et al., 2011). Results in the literature are not scarce, but they are mostly focused on prisons, and their endpoint is usually just risk of occurrence (as opposed to time to occurrence, or equivalently its cumulative hazard). Many results are based on not very recent and/or local data. Several studies suggest that jail/prison stays

Received: January 24, 2025. **Revised:** August 7, 2025. **Accepted:** August 25, 2025

may have complex effects on recidivism. For example, analysing data from Florida, Mears et al. (2016) found that recidivism initially even increases with longer sentences, it decreases shortly after, and the association vanishes over time. Durose et al. (2014) reported that 67.8% of 404,638 prisoners released in 2005 in 30 states were arrested within 3 years of release, and 76.6% were arrested within 5 years of release. Among prisoners released in 2005 in 23 states 49.7% had a parole, probation violation, or arrest for a new offense in 3 years; and 55.1% had a parole, probation violation, or arrest in 5 years. Western et al. (2021) focus on New York jails, finding a strong association of risk of jail recidivism with race, and, as mentioned above, generally very high rates.

Our aim is to enrich the literature with a comprehensive analysis of the recent U.S. jailed population. Data were provided by the NYU Public Safety Lab's Jail Data Initiative (https://jaildatainitiative.org), which collects individual-level jail data from over a thousand sources throughout the U.S. Our dataset includes 686,319 observations, encompassing 550,994 individuals who have been incarcerated at least once in the period 2020–2023, or were already in jail at the beginning of 2020. To our knowledge, this is the first study to focus on post-release freedom duration in addition to the likelihood of occurrence of a new detention; and it is currently the study using the most recent and largest individual-based data set.

In order to analyse our data we make a contribution to the extant literature on multi-state models (see Meira-Machado et al., 2009 for a review). Indeed in our data individuals occupy one of two states: in detention, or free; none of which is absorbing. There are therefore two possible transitions: freedom to jail, and jail to freedom. Our framework will be based on quantile regression (Koenker & Bassett, 1978). Quantile regression methods for multi-state models can be found in Beyersmann and Schumacher (2008) and Farcomeni and Geraci (2020). The literature on quantile regression for time-to-event data is very rich, and multi-state models encompass recurrent event (Sun et al., 2016), competing risk (Peng & Fine, 2009; Qu et al., 2023), and semi-competing risk (Hsieh & Wang, 2018; Li & Peng, 2011) models. Our methodological contribution is three fold. First of all, we introduce a cure fraction (Peng & Yu, 2021) in the multi-state quantile regression model. This implies that a subset of the subjects are not expected to ever experience the transition of interest. In our context, it will be used to significantly improve goodness of fit in the face of a low event rate (Jackson et al., 2022). The use of a cure fraction in multi-state models has been introduced in Conlon et al. (2014) in the classical framework, and we will discuss it here for the first time in the quantile regression one. Our second contribution involves the introduction of a discrete latent variable to capture unobserved heterogeneity, and cluster subjects. We will propose an infinite mixture model, thereby not having to specify in advance the support of the latent variable. This assumption leads to a nonparametric Bayesian approach, specifically, to specification of a Dirichlet process prior for the mixing distribution. Our third contribution will be the development and implementation of a slice sampling approach for the resulting Dirichlet process mixture (Escobar & West, 1995; MacEachern & Müller, 1998), and of a batched Markov Chain Monte Carlo (MCMC) strategy that can be used for large data sets (like the one that motivates our work).

The rest of the article is as follows: in the next section, we introduce the latent class multi-state quantile regression model with a cured fraction. In Section 3, we describe the Bayesian inferential procedure. The scalable MCMC strategy is described in Section 3.3. A simulation study is reported in Section 4, and the analysis of our motivating example is reported in Section 5. Conclusions are given in Section 6. R codes are available as online supplementary material.

2 Multi-state quantile regression with unobserved heterogeneity and a cured fraction

2.1 Background on multi-state quantile regression

Consider a stochastic process S_t, for $t \geq 0$, that evolves in continuous time and discrete state-space $\{1, \ldots, s\}$, where $s \geq 2$. Under mild assumptions (made explicit below) it is possible to characterize the stochastic process S_t in terms of its first hitting times. Assume that the state m is not an absorbing state. Let $T_m = \inf\{t \geq 0 : S_t \neq m, S_0 = m\}$ be the random variable representing the time of the first jump from state m. Additionally, let Y_j denote the sequence of states visited by the

continuous-time Markov chain (regardless of dwelling times), with Y_0 as the initial state. Under Markov homogeneity assumptions, the stochastic process is entirely determined by

$$F_{mj}(t) = \Pr(T_m \leq t,\, Y_1 = j \mid Y_0 = m) = \Pr(T_{mj} \leq t), \quad t > 0 \tag{1}$$

where T_{mj} is the time taken to transition from state m to state j, $m = 1, \ldots, s$, and $j \neq m$. Farcomeni and Geraci (2020) propose a quantile regression model for multi-state data. They defined the variable T^*_{mj} as $T^*_{mj} = T_{mj}$ if the transition is to state j, and $T^*_{mj} = \infty$ if the transition is to some other state. When the follow-up is closed in state m, there is an independent censoring time C_m, where $C_m = \infty$ if a transition occurs. By defining the $p \times 1$ vector of predictors $X = (X_1, X_2, \ldots, X_p)'$, the conditional τ-quantile function can be defined as

$$Q_{mj}(\tau \mid X) = \inf\{t : \Pr(T^*_{mj} \leq t,\, Y_1 = j \mid X,\, Y_0 = m) \geq \tau\}. \tag{2}$$

The function $Q_{mj}(\tau \mid X)$ is the left inverse of the conditional cumulative distribution function (CDF) $\Pr(T^*_{mj} \leq t,\, Y_1 = j \mid X,\, Y_0 = m)$. To explicitly describe the relationship between $Q_{mj}(\tau \mid X)$ and X for a given τ, one can specify a transformation model of the form

$$Q_{mj}(\tau \mid X) = g_{mj}\{\beta_0^{(mj)}(\tau) + X'\beta_{mj}(\tau)\}, \tag{3}$$

where $g_{mj}(\cdot)$ represents a known monotone link function, $\beta_0^{(mj)}(\tau)$ denotes a quantile-specific intercept, and $\beta_{mj}(\tau) = (\beta_1^{(mj)}(\tau), \beta_2^{(mj)}(\tau), \ldots, \beta_p^{(mj)}(\tau))'$ a vector of quantile-specific regression coefficients. The kth element of $\beta_{mj}(\tau)$ can be understood as the variation in the τth quantile of T^*_{mj}, considering the scale of $g_{mj}(\cdot)$, when X_k increases by one unit while keeping all other predictors constant.

The use of a link function $g_{mj}(\cdot)$ is in parallel with classical formulations for generalized linear models. In this context, a link function is mostly required to take care of the support of the response variable (Bottai et al., 2010) and to make the linearity assumption more credible. See also Mu and He (2007) and Geraci and Jones (2015). Since we are modelling times, which are bounded from below, a logarithmic link function seems to be the most natural, but any monotone function mapping $\mathcal{R}^+$ to $\mathcal{R}$ is suitable.

2.2 Cured fraction in multi-state quantile regression

Inspired by the approach of Conlon et al. (2014), who introduced a multi-state Markov model with an incorporated cured fraction to jointly model recurrence and death in colon cancer, we now describe a multi-state quantile regression model with a cured fraction. The unobserved cured status identifies individuals who are bound not to experience the transition of interest. A latent indicator will allow us to distinguish between individuals who are recurrent, and individuals who instead remain within the state for an infinite time. Through the specification of an appropriate generalized linear model we can assess the effect of covariates on the probability of being cured towards the transition of interest.

From a practical point of view, our data exclusively includes individuals who have been in jail at least once. The cure indicator for the transition from freedom to jail will differentiate between recurring subjects—those who after being released will eventually return to jail—and subjects who will not experience a new entrance into jail. However, this event will be difficult to interpret, as it will inevitably include both subjects who will remain in the 'freedom' state, and subjects who will enter a prison (instead of a jail) in the future. The advantage of the cure fraction in our example will reside mostly in a considerably better fit.

Let us consider a generic multi-state model with discrete state-space $\{1, \ldots, s\}$ where $s \geq 2$, as defined in Section 2. In order to relax the common assumption that all subjects must eventually experience the event (e.g. that there exist an event time for each subject, which is only possibly censored), let us introduce a variable $R^{(mj)}$, with $R^{(mj)} = 1$ if the subject is cured for the transitions from the state m to state j, and 0 otherwise; such that, for all $t \geq 0$,

$$\Pr(T_{mj} \leq t \mid R^{(mj)} = 1) = 0. \tag{4}$$

In words, a fraction of the population is allowed to never experience the transition from state m to state j. On the other hand, whenever the latent variable $R^{(mj)} = 0$ individuals will experience the event (even if the exact event time is not observed due to censoring). In practical terms, the binary latent variable $R^{(mj)}$ identifies two sub-populations within our sample. We model

$$P(R^{(mj)} = 1 \mid Z) = \eta_{mj}\left(\gamma_0^{(mj)} + Z'\gamma_{mj}\right), \tag{5}$$

where Z represents a vector of covariates, $\eta_{mj}(\cdot)$ is a link function for binary regression models, $\gamma_0^{(mj)}$ is an intercept, and γ_{mj} is a parameter vector explaining the effect of Z on the probability of being cured.

This latent structure allows us to model the individual-specific probability of being cured with respect to the transition from state m to state j, denoted by $P(R^{(mj)} = 1 \mid Z)$. If an individual is cured, that is, if $R^{(mj)} = 1$, the probability of experiencing the transition is zero. This formulation is commonly used to relax the standard assumption that all individuals will eventually experience the event, as typically imposed in classical Cox models. The conditional probability being zero is essential to properly define the subgroup of cured individuals (also referred to as long-term survivors). If the conditional probability were instead strictly positive, the model would correspond to a standard latent class framework, where individuals differ in their unobserved risk levels (e.g. median time-to-event) but are still all susceptible to the event. This structure is in fact incorporated in our Dirichlet Process Mixtures (DPM)-based approach, which allows for heterogeneous time-to-event distributions among the non-cured individuals. In summary, our model accommodates both zero and non-zero conditional probabilities, and explicitly modelling a cured component with a point mass at infinity proves crucial for improving model fit and interpretability.

We can derive

$$\begin{aligned} \Pr\left(T_{mj} \leq t, R^{(mj)} \mid Z\right) &= \sum_{a=0}^{1} \Pr(T_{mj} \leq t \mid R^{(mj)} = a) \Pr(R^{(mj)} = a \mid Z) \\ &= \Pr(T_{mj} \leq t \mid R^{(mj)} = 0) \Pr(R^{(mj)} = 0 \mid Z). \end{aligned} \tag{6}$$

The conditional quantile function (2) can then be expressed as

$$Q_{mj}(\tau \mid X) = \inf\{t : \Pr(T^*_{mj} \leq t, Y_1 = j \mid R^{(mj)} = 0, X, Y_0 = m) \geq \tau\}, \tag{7}$$

with $\tau \in (0, 1)$; where we still parameterize the expression above as a function of covariates as in (3).

Note that since the model is conditional on the covariates there can be overlap between X and Z. In fact, in our analysis, we will specify the same covariates for the cure fraction and for the duration models.

2.3 Latent class multi-state quantile regression

The model discussed in the previous section assumes that all heterogeneity is captured by the covariate vectors X and Z. In our context, this is not credible, as several important determinants of spell length and probability of (lack of) recidivism are not observed, possibly not even measurable. We therefore allow for the possibility of unobserved heterogeneity in the model by including a discrete latent variable K. This is convenient for two reasons. First of all, a discrete latent variable can adapt to almost any distribution for the resulting random effect, therefore not requiring the use of specific parametric assumptions (e.g. that random intercepts are Gaussian). Secondly, a by-product of the model is a classification of trajectories into homogeneous classes, which can allow us to identify interpretable sub-populations of subjects.

Formally, we let $\Pr(K = k) = w_k$, with $w_k > 0$ and $\sum_w w_k = 1$, for the discrete latent variable; and modify the model equations as

$$\begin{cases} \Pr(R^{(mj)} = 1 \mid Z, K = k) = \eta_{mj}\left(\gamma_{0k}^{(mj)} + Z'\gamma_{mj}\right) \\ Q_{mj}(\tau \mid X, K = k) = g_{mj}\{\beta_{0k}^{(mj)}(\tau) + X'\beta_{mj}(\tau)\}, \end{cases}$$

where

$$Q_{mj}(\tau \mid X, k) = \inf\{t : \Pr(T^*_{mj} \leq t, Y_1 = j \mid R^{(mj)} = 0, X, Y_0 = m, k) \geq \tau\}, \tag{8}$$

In words, intercepts are now subject-specific. Subjects belonging to the same latent class share a common value for the intercept, which can be interpreted as a common propensity to transition from state m to state j, or to stay in state m indefinitely, conditionally on the covariates. It should be noted that it would be straightforward to relax the assumption that the regression coefficients β_{mj} and γ_{mj} are independent of K, but at the price of an explosion in the number of parameters. The improved fit would also definitely come at the price of interpretability of the parameter estimates.

2.4 Cumulative hazard

Classical multi-state models involve the estimation of a hazard function, and consequently, the cumulative hazard $H(t)$. The latter is a direct estimate of the number of events that are expected at each time interval. A limitation of classical methods is that they are not applicable in big data contexts, as is the case of our motivating data set.

Given that we are interested in jail recidivism, an estimate of cumulative hazard is important in our application, as $H(t)$ corresponds to the expected number of transitions in the interval $(0, t)$. Here, we describe how to obtain it from our model estimates. First of all, recall that

$$\hat{H}(t) = \int_0^t -\frac{\partial \log(\hat{S}(v))}{\partial v}\, \mathrm{d}v, \tag{9}$$

where $\hat{S}(v)$ is a short-hand notation for any (conditional) survival function of transition times.

In our case, we will be able to estimate conditional cumulative hazard functions, possibly also conditionally on the discrete latent variable, depending on which estimated survival function is plugged into (9).

As an example, consider $Q_{mj}(\tau \mid X, k)$ as defined in (8), which corresponds to the τ quantile of the time to transition from state m to state j, conditional on $R^{(mj)} = 0$, latent state k, covariate configuration X.

Our modelling approach gives only point estimates for quantile, and hence survival, functions. Thus, we need to estimate our model on a grid of different quantiles if we wish to invert the quantile function to obtain (9). To fix the ideas, assume that we estimate $Q_{mj}(\tau \mid X, k)$ at $\tau_1 < \tau_2 < \cdots < \tau_B$. In our application, we used $B = 100$ equally spaced values, which is feasible given that (batched) MCMC sampling is relatively fast and can proceed in parallel.

The estimated quantile function is then inverted to obtain a survival function evaluated at B time points. One can in fact compute

$$\hat{S}_{mj}(t \mid X, k) = \inf\{\tau : \hat{Q}_{mj}(\tau \mid X, k) \leq t\}$$

and plug it in (9) to obtain $\hat{H}_{mj}(t \mid X, k)$.

In (9), the first derivative of $\log(\hat{S}_{mj}(t \mid X, k))$ is simply approximated by the first differences in the logarithm of the survival function.

A similar approach is used to obtain the cumulative hazard after marginalization with respect to $R^{(mj)}$.

3 Inference

In the frequentist framework, quantile regression estimates are obtained as any solution to a minimization problem based on the quantile check function $\rho_\tau(u) = u(\tau - \mathbb{1}(\mathbb{u} \leq \mathbb{0})$, with $\mathbb{1}(\cdot)$ denoting the indicator function. Yu and Moyeed (2001) show that minimization of the loss function is equivalent to the maximization of a working likelihood function based on the asymmetric Laplace distribution (ALD):

$$f(t \mid \beta, \sigma, \tau, X) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-(\tau - \mathbb{1}(t \leq X'\beta))(t - X'\beta)\sigma^{-1}\right\}. \tag{10}$$

This paves the way for a likelihood-based Bayesian approach, facilitating the implementation of MCMC methods.

3.1 Bayesian inference for multi-state quantile regression with a cured fraction

We first introduce the inferential approach for the case without latent classes (i.e. with K having a single support point). Assume to have a sample of N subjects. For $i = 1, \ldots, N$, let $\mathbf{t}_i^{(mj)} = (t_{i1}^{(mj)}, \ldots, t_{in_i}^{(mj)})$ denote the set of observed transition times between the states m and j for the ith subject, with n_i representing the total number of transitions from m to j observed for the individual i. Let $T^{(mj)} = (\mathbf{t}_1^{(mj)}, \ldots, \mathbf{t}_N^{(mj)})$ be the set containing all the observed transition times from m to j across the N observed individuals. Let $c_{il}^{(mj)} = 1$ if the lth observation of the ith individual is censored, and $c_{il}^{(mj)} = 0$ if the observation is uncensored. For each unit we have observed covariates $X_i^{(mj)} = (X_{i1}^{(mj)}, \ldots, X_{in_i}^{(mj)})$ and $Z_i^{(mj)}$. We can write the likelihood of the model as:

$$\mathcal{L}(\beta_{mj}, \sigma_{mj}, \gamma_{mj}) = \prod_{i=1}^{N} \left(\Pr\left(R_i^{(mj)} = 0 \mid Z_i^{(mj)}, \gamma_{mj}\right) \prod_{l=1}^{n_i} f\left(t_{il}^{(mj)}, c_{il}^{(mj)} \mid \beta_{mj}, \sigma_{mj}, X_{il}^{(mj)}\right) \right), \tag{11}$$

which in practice is obtained by multiplying, for each observed statistical unit, the probability of being cured by the product of the pdfs of the transition times. An explicit expression is obtained as

$$\begin{aligned} & f(t_i, c_i \mid \beta_{mj}, \sigma_{mj}, \gamma_{mj}, X_i, Z_i) \\ & = \prod_{i:\, n_i = 1} \left\{ \left[\left(1 - \eta_{mj}\left(Z_i^{(mj)'} \gamma_{mj}\right)\right) f\left(t_{i1}^{(mj)} \mid \beta_{mj}, \sigma_{mj}, \tau, X_{i1}^{(mj)}\right) \right]^{\left(1 - c_{i1}^{(mj)}\right)} \right. \\ & \left. \left[\eta_{mj}\left(Z_i^{(mj)'} \gamma_{mj}\right) + \left(1 - \eta_{mj}\left(Z_i^{(mj)'} \gamma_{mj}\right)\right) \left(1 - F\left(t_{i1}^{(mj)} \mid \beta_{mj}, \sigma_{mj}, \tau, X_{i1}^{(mj)}\right)\right) \right]^{c_{mj}^{(i1)}} \right\} \\ & \prod_{i:\, n_i > 1} \left(1 - \eta_{mj}\left(Z_i^{(mj)'} \gamma_{mj}\right)\right) \prod_{l=1}^{n_i} \left\{ f\left(t_{il}^{(mj)} \mid \beta_{mj}, \sigma_{mj}, \tau, X_{il}^{(mj)}\right)^{\left(1 - c_{il}^{(mj)}\right)} \right. \\ & \left. \left(1 - F\left(t_{il}^{(mj)} \mid \beta_{mj}, \sigma_{mj}, \tau, X_{il}^{(mj)}\right)\right)^{c_{il}^{(mj)}} \right\}, \end{aligned} \tag{12}$$

where $f(\cdot)$ denotes the ALD density defined in (10), and $F(\cdot)$ denotes its CDF

$$\begin{aligned} F(t \mid \beta, \sigma, \tau, X) &= \int f(t \mid X, \beta, \tau)\, \mathrm{d}t \\ &= \mathbb{1}(t > X'\beta) - (\mathbb{1}(t > X'\beta) - \tau) \exp\left\{ (\mathbb{1}(t \le X'\beta) - \tau)(t - X'\beta)\sigma^{-1} \right\}. \end{aligned} \tag{13}$$

As customary in the Bayesian literature, prior distributions should summarize available information if possible. Here, we make a simple suggestion for a reasonable default prior scheme. We use a prior independence assumption to specify a different distribution separately for each parameter. Since the scale parameters σ_{mj} must be positive, we set as prior a Gamma distribution, centred on the unity, with large variance. For the regression parameter vectors β_{mj} and γ_{mj}, we specify zero-centred Gaussian distributions.

By applying Bayes theorem, the posterior density can be written as

$$\begin{aligned} \pi(\beta_{mj}, \sigma_{mj}, \gamma_{mj} \mid T^{(mj)}, X, Z, c^{(mj)}) &\propto J(\beta_{mj}, \sigma_{mj}, \gamma_{mj} \mid T^{(mj)}, X, Z, c^{(mj)}) \\ &= \pi(\beta_{mj})\pi(\sigma_{mj})\pi(\gamma_{mj}) \prod_{i=1}^{N} f(\mathbf{t}_i \mid c^{(mj)}, \beta_{mj}, \sigma_{mj}, \gamma_{mj}, X_i, Z_i). \end{aligned} \tag{14}$$

In order to approximate the posterior, we use an MCMC approach. Our MCMC scheme is fast as the observed likelihood can be computed explicitly, without the need for augmentation (i.e. sampling the latent cure indicator).

We sample from the posterior distribution by using a random walk Metropolis algorithm with a multivariate Gaussian proposal density. Adaptive schemes ensure proper exploration of parameter space, and promote rapid convergence of MCMC (Rosenthal, 2011). At each iteration we sample $(\beta^*_{mj}, \sigma^*_{mj}, \gamma^*_{mj})$ from $\mathcal{N}((\beta^{(h)}_{mj}, \log(\sigma^{(h)}_{mj}), \gamma^{(h)}_{mj}), \Sigma)$, where $(\beta^{(h)}_{mj}, \log(\sigma^{(h)}_{mj}), \gamma^{(h)}_{mj})$ is the vector of last accepted values and Σ represents the covariance matrix of the proposal. The proposal is accepted with probability:

$$\min\left(1, \frac{J(\beta^*_{mj}, \log(\sigma^*_{mj}), \gamma^*_{mj} \mid T^{(mj)}, X, Z, c^{(mj)})}{J(\beta^{(h)}_{mj}, \log(\sigma^{(h)}_{mj}), \gamma^{(h)}_{mj} \mid T^{(mj)}, X, Z, c^{(mj)})}\right), \tag{15}$$

with $J(\cdot)$ corresponding to the numerator of the posterior distribution as in the right hand side of (14). We implement the MCMC algorithm for a sufficiently large number of iterations; and we assess convergence by examining trace and autocorrelation function plots.

Although the Asymmetric Laplace (AL) distribution is adopted as a working likelihood, credible intervals derived directly from MCMC samples may be unreliable, since the AL does not represent the true data-generating process. To mitigate this, Yang et al. (2016) introduces a variance correction that ensures asymptotically valid inference. Let $\hat{\Sigma}$ denote the MCMC estimate of the covariance matrix for the parameters. The corrected covariance matrix is obtained as

$$\frac{\tau(1-\tau)}{\hat{\sigma}} \hat{\Sigma} \sum_i X_i^{(mj)} X_i^{(mj)'} \hat{\Sigma}.$$

It shall be noted that Yang et al. (2016) use a fixed σ, while in our work we estimate σ. The correction is still seen to be useful, as testified by the simulation reported below.

3.2 Dirichlet process mixtures of multi-state quantile regression models

We now focus on the infinite mixture of quantile regression models defined in Section 2.3, that is, to the case where the latent variable K has more than one support point. The estimation procedure relies on a relatively complex algorithm, structured into two main phases, as commonly employed in Bayesian nonparametric mixture models. First, the random partition is updated via slice sampling (Kalli et al., 2011; Walker, 2007), as detailed in Section 3.2.1. This step determines the number of mixture components and samples the weights w_k. Then, the model parameters are updated through the random walk Metropolis algorithm, as described in detail in the previous subsection for the multi-state quantile regression with cured fraction. The parameters modulated by the latent variable, namely the intercepts $\beta^{(mj)}_{0k}$ and $\gamma^{(mj)}_{0k}$, are updated conditionally on the clustering structure.

In order to simplify the notation, we will not explicitly specify the transition mj when it is clear from the context. The density of the kth mixture component can be obtained by distinguishing between two cases: if $n_i = 1$, then we have:

$$\begin{aligned} f(t_i, c_i \mid K=k, \theta, X, Z, c) = \Big\{ & \left(1 - \Pr\left(R^{(mj)} = 1 \mid \gamma^{(mj)}_{0k}, \gamma_{mj}, Z_i^{(mj)}\right)\right) \\ & f\left(t_i \mid \beta^{(mj)}_{0k}, \beta_{mj}, \sigma_{mj}, X_i^{(mj)}\right) \Big\}^{\left(1-c^{(mj)}_{i1}\right)} \Big\{ \Pr\left(R^{(mj)} = 1 \mid \gamma^{(mj)}_{0k}, \gamma_{mj}, Z_i^{(mj)}\right) \\ & + \left(1 - \Pr\left(R^{(mj)} = 1 \mid \gamma^{(mj)}_{0k}, \gamma_{mj}, Z_i^{(mj)}\right)\right)\left(1 - F\left(t_i^{(mj)} \mid \beta^{(mj)}_{0k}, \beta_{mj}, \sigma_{mj}, X_i^{(mj)}\right)\right) \Big\}^{c^{(mj)}_{i1}}, \end{aligned} \tag{16}$$

where θ is a short notation for model parameters. If $n_i > 1$, we have:

$$f(t_i, c_i \mid K=k, \theta, X, Z, c) = \left(1 - \Pr\left(R^{(mj)} = 1 \mid \gamma_{0k}, \gamma_{mj}, Z_i^{(mj)}\right)\right) \prod_{l=1}^{n_i} \left\{ f\left(t_{il} \mid \beta_{0k}, \beta_{mj}, \sigma_{mj}, X_{il}^{(mj)}\right)^{\left(1-c_{il}^{(mj)}\right)} \left(1 - F\left(t_{il} \mid \beta_{0k}, \beta_{mj}, \sigma_{mj}, X_{il}^{(mj)}\right)\right)^{c_{il}^{(mj)}} \right\}, \tag{17}$$

where, clearly,

$$f(t \mid \beta_{0k}, \beta, \sigma, \tau, X) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-(\tau - \mathbb{1}(\mathbb{t} \leq (\beta_{0\mathbb{k}} + \mathbb{X}'\beta)))(\mathbb{t} - (\beta_{0\mathbb{k}} + \mathbb{X}'\beta))\sigma^{-1}\right\}$$

and

$$\Pr\left(R^{(mj)} = 1 \mid \gamma_{0k}, \gamma, Z\right) = \eta\left(\gamma_{0k} + Z'\gamma\right).$$

By specifying the Dirichlet process with centring measure G_0 and scaling parameter $M > 0$ as prior on the mixing distribution, we get a Dirichlet process mixture. DPM is a class of statistical models suitable for clustering and density estimation. Dirichlet process mixtures provide a way to model the uncertainty about the number of components, automatically determining the appropriate number of clusters.

In this formulation, we have a regression equation structure with intercepts dependent on a discrete latent variable; thus, the prior setting for the parameters independent of the latent classes is the same as defined in Section 3.1, while the latent intercepts, i.e. $\alpha_k = (\beta_{0k}, \gamma_{0k})$, will be assumed to arise from a bivariate Gaussian prior centred at zero. Therefore, we fix $G_0 = N_2(0, \nu I_2)$, with I_2 denoting a diagonal matrix.

This approach does not require a prior specification for the support of the latent variable K, not even an upper bound, and can be represented as

$$f(t_i, c_i \mid \theta, X, Z) = \sum_{k=0}^{\infty} w_k f(t_i, c_i \mid K = k, \theta, X, Z), \tag{18}$$

where $w_k \geq 0$ are mixture weights such that $\sum_{k=0}^{\infty} w_k = 1$. We may also further rewrite the model (18) in a hierarchical form:

$$\begin{aligned} t_i \mid \alpha_k &\overset{ind}{\sim} p_{\alpha_k} \\ \alpha_k \mid G &\overset{ind}{\sim} G \\ G &\sim \mathrm{DP}(MG_0). \end{aligned} \tag{19}$$

For discussions on how to handle the posterior, see, for example, Wade and Ghahramani (2018) and Lavigne and Liverani (2024), and references therein. The relationship between w_k and $G(\cdot)$ will be made clear in the next subsection.

3.2.1 *Slice sampling*

Walker (2007) and Kalli et al. (2011) proposed a slice sampling scheme for DPM. This approach efficiently samples the weights of the components by defining slices based on the current weights, facilitating the exploration of the parameter space while ensuring that the samples accurately represent the underlying distribution. Specifically, this procedure involves the introduction of uniform auxiliary variables, which render the mixture model conditionally finite. Let us consider the model defined in (18) and construct the mixing measure G using the stick-breaking representation, that is

$$G(\cdot) = \sum_{k=0}^{\infty} w_k \delta_{\alpha_k}(\cdot), \tag{20}$$

where $w_k = v_k \prod_{j=1}^{k-1}(1 - v_j)$, with $v_k \sim \text{Be}(1, M)$ independently, $\delta_{\alpha_k}(\cdot)$ denotes the Dirac measure giving mass 1 at α_k and 0 elsewhere, and $\alpha_k \sim G_0$ independently. Let w and α represent the sequences of w_k and α_k, respectively. Using this representation, we can explicitly express the DPM model as:

$$f(t_i, c_i \mid \theta, X_i, Z_i) = \sum_{k=0}^{\infty} w_k f(t_i, c_i \mid K = k, \theta, X_i, Z_i).$$

However, since this is still an infinite sum, it poses challenges for computation and posterior simulation. To address this, we introduce latent variables u_i, transforming the infinite sum into a finite one:

$$f(t_i, c_i, u_i \mid \theta, X_i, Z_i) = \sum_{k=0}^{\infty} \mathbb{1}(u_i < w_k) f(t_i, c_i \mid K = k, \theta, X_i, Z_i).$$

By integrating over u_i, we recover the original expression.

This sum involves only a finite number of terms, as we now outline. Define $C_w(u) = \{k \mid u < w_k\}$. In practice, we need to consider k up to $k^* = \min\{k \mid u_i > 1 - \sum_{j=1}^{k} w_j\}$, because for all $k > k^*$, $w_k < 1 - \sum_{j=1}^{k^*} w_j < u_i$. Therefore, the sum is finite.

Consider now d_i, with $f(d_i = k \mid u_i, w) \propto \mathbb{1}(u_i < w_k)$, which is uniform over $C_w(u_i)$. The joint density becomes:

$$f(t_i, c_i, u_i, d_i \mid \theta, X_i, Z_i) = \mathbb{1}(u_i < w_{d_i}) f(t_i, c_i \mid K = k, \theta, X_i, Z_i).$$

By integrating over d_i and then over u_i, we recover the previous expressions. This formulation allows for straightforward implementation of a Gibbs sampler for posterior simulation. In the augmented model, the joint distribution is:

$$f(t_i, c_i, u_i, d_i \mid \theta, X_i, Z_i) = \prod_{i=1}^{N} \mathbb{1}(u_i < w_{d_i}) f(t_i, c_i \mid K = k, \theta, X_i, Z_i).$$

We can construct a Gibbs sampler that updates w_k, α_k, u_i, and d_i by sampling from their full conditional posterior distributions. Since $w_k = v_k \prod_{j=1}^{k-1}(1 - v_j)$, with $v_k \sim \text{Be}(1, M)$, we define $A_k = \{i \mid d_i = k\}$ and $B_k = \{i \mid d_i > k\}$. Then,

$$f(v_k \mid d) = \text{Be}(1 + |A_k|, M + |B_k|), \tag{21}$$

subject to the condition that $u_i < w_{d_i}$. Similarly,

$$f(\alpha_k \mid \theta, T, X, Z, c, d) \propto G_0(\alpha_k) \prod_{i \in A_k} f(t_i, c_i \mid K = k, \theta, X_i, Z_i), \tag{22}$$

where $\alpha_k = (\beta_{0k}, \gamma_{0k})$ is updated, using a Metropolis step, only if A_k is non-empty. For empty indices, α_k can be sampled from the prior G_0. Then, we update the posterior distributions of $\theta = (\beta, \gamma, \sigma)$ with another Metropolis step sampling from

$$f(\theta \mid \alpha, T, X, Z, c, d) \propto f(\theta) \prod_{i=1}^{N} f(t_i, c_i \mid K = d_i, \theta, X_i, Z_i). \tag{23}$$

Finally, the full conditional for $f(u_i \mid d_i, w) = \text{Unif}(0, w_{d_i})$, while d_i, conditionally on u, is a discrete uniform distribution on the set $C_w(u)$.

3.3 Dealing with big data: partial batch MCMC

The proposed method scales well up to about two hundred thousand events. For larger sample sizes, it might become excessively computationally intensive. In our sample, we have about seven hundred thousand transitions for slightly more than half a million individuals.

In order to address this issue, we derive a batch MCMC scheme (Wu et al., 2022). To our knowledge, the partial approach we propose is novel and, as we now argue and illustrate, simple and effective. We propose to use the entire data set at the clustering phase and mixture weights update of the MCMC sampler; while we use subsamples to update the posterior distributions of the kernel parameters. Specifically, we obtain a partition of the data in $\Omega > 1$ blocks. At each iteration, we use only one of the blocks to update the kernel parameters. After Ω iterations, an epoch ends, and a new partition is generated. We repeat until the desired number of MCMC iterations. Batching speeds up each individual MCMC update, at the cost of increased variability in the sampled parameters. However, random sampling guarantees that the first posterior moment is approximated correctly. The second posterior moment can be easily corrected through a straightforward application of the Bernstein-von Mises theorem. In fact, as the number of observations in each batch increases, it can be shown that the posterior variance of each parameter is approximately Ω times the posterior variance that would be obtained if the entire sample were used in each MCMC iteration. This is illustrated in a brief simulation study below, where we check the validity of this proposal.

4 Simulation study

In this section, we describe a simulation study. We first focus on a model that does only involve a cured fraction, then we allow for different support points for the latent variable. Finally, we empirically verify the effectiveness of the proposed batch sampling scheme and covariance correction.

4.1 Multi-state quantile regression with a cured fraction

The aim of this first simulation study is to show that, even in the absence of latent classes, when sub-populations of cured individuals exist with respect to a specific transition, the estimates obtained using the model proposed by Farcomeni and Geraci (2020) are less precise than those from a Bayesian model with a cured fraction. To this end, we simulate a multi-state process for a generic transition. We allow for two sample sizes, $N = 1{,}000$ and $N = 10{,}000$. Two continuous covariates are generated independently from $\mathcal{U}(-1, 1)$ and $\mathcal{N}(0, 1)$. The regression parameters for transition times are fixed as $\beta_0 = 0.2$, $\beta_1 = 0.3$, and $\beta_2 = -0.2$. For the cured fraction, we use a probit link and only the first covariate, and fix $\gamma_0 = 0.2$ and $\gamma_1 = 0.6$. As in our proposed modelling scheme, we parameterize the latent variable R_i through a probit transform of the linear predictor based on the γ parameters. Whenever $R_i = 0$, we sample log holding times from a Gaussian random variable with expected value corresponding to the linear predictor based on β parameters, and unit variance. An independent censoring time is also simulated. The censoring time is drawn from a Uniform distribution over the interval (0, 100): if it occurs before the event time, the observation is treated as censored and the corresponding transition is not recorded.

After we generate the data, we estimate our Bayesian multi-state quantile regression with cured fraction, and classical multi-state quantile regression (without a cured fraction). We estimate each model at quantile levels $\tau = (0.25, 0.5, 0.75)$. We replicate data generation and estimation 100 times. Once the estimates are obtained, we evaluate the performance in terms of root mean squared error (RMSE) for parameter estimates, and in terms of relative root mean squared error (RRMSE), that is RMSE divided by the true value, for predicted quantiles.

In Tables 1 and 2, we compare the results obtained with the implementation of the proposed Bayesian MSM with (MSQR-CF) and without (MSQR) a cured fraction.

As expected, MSQR seems to be biased in these scenarios. and the bias does not decrease with the sample size. This is related to the fact that cured individuals have (censored) large times, therefore misspecification affects estimates at $\tau = 0.75$ much more than at $\tau = 0.25$. However, the MSQR model shows a larger MSE than MSQR-CF nonetheless also at $\tau = 0.25$. The same happens if one varies the true data generating distribution (not shown for reasons of space). While the RMSE for parameter estimates is in general rather large, it can be expected to decrease

Table 1. Simulation study

	MSQR-CF					MSQR		
	β_0	β_1	β_2	γ_0	γ_1	β_0	β_1	β_2
$\tau = 0.25$								
$N = 1{,}000$	0.020	0.139	0.098	0.059	0.062	0.097	0.102	0.077
$N = 10{,}000$	0.009	0.139	0.097	0.031	0.022	0.087	0.099	0.077
$\tau = 0.50$								
$N = 1{,}000$	0.044	0.060	0.034	0.056	0.061	0.343	0.218	0.101
$N = 10{,}000$	0.019	0.022	0.011	0.020	0.019	0.315	0.173	0.085
$\tau = 0.75$								
$N = 1{,}000$	0.102	0.338	0.188	0.054	0.062	1.864	1.410	0.745
$N = 10{,}000$	0.065	0.299	0.185	0.018	0.018	1.600	1.190	0.615

Note. Root mean squared error (RMSE) of the posterior means across 100 samples for Bayesian multi-state quantile regression with cured fraction (MSQR-CF) and multi-state quantile regression (MSQR); at different quantiles τ and sample sizes N.

Table 2. Simulation study

	MSQR-CF	MSQR
$\tau = 0.25$		
N = 1,000	0.0253	1.3967
N = 10,000	0.0247	1.3837
$\tau = 0.50$		
N = 1,000	0.0235	1.3786
N = 10,000	0.0234	1.3663
$\tau = 0.75$		
N = 1,000	0.0458	1.3347
N = 10,000	0.0449	1.3239

Note. Relative Root Mean Squared Error (RRMSE) of estimated conditional quantiles across 100 samples for Bayesian multi-state quantile regression with cured fraction (MSQR-CF), and multi-state quantile regression (MSQR); at different quantiles τ and sample sizes N.

substantially at larger sample sizes (like in our real data example). Furthermore, the RRMSE for the predicted conditional quantiles is more than acceptable.

4.2 Latent class multi-state quantile regression with cured fraction

In this second simulation framework, we also introduce unobserved heterogeneity across individuals. The aim of this study is to show that the Dirichlet process mixture of multi-state quantile regression model with cured fraction is well suited to handle such scenarios, outperforming the other models considered. We simulate data from a latent class model. Once again we let $N = 1{,}000$ or $N = 10{,}000$. We fix the covariate parameters for transition times as $\beta_1 = 0.7$ and $\beta_2 = -0.45$. For the cure fraction, we use a probit link with $\gamma_1 = 0.3$. We allow for two latent classes, where $\beta_{01} = \log(10)$, $\beta_{02} = \log(2)$, $\gamma_{01} = 0.4$, and $\gamma_{02} = -0.4$. We generate data by assigning each individual to a latent class with equal probability. Conditioned on the assigned group, we proceed as in the previous subsection.

We evaluate the efficacy of the latent class model in comparison to the marginal model proposed by us, and, as before, multi-state quantile regression without a cure fraction. Each model is estimated across quantile levels $\tau = (0.25, 0.5, 0.75)$. Results are in Tables 3 and 4.

Table 3. Simulation study

	DPM-MSQR-CF			MSQR-CF			MSQR	
	β_1	β_2	γ_1	β_1	β_2	γ_1	β_1	β_2
$\tau = 0.25$								
$N = 1{,}000$	0.205	0.071	0.077	0.200	0.055	0.100	0.352	0.114
$N = 10{,}000$	0.071	0.045	0.055	0.084	0.000	0.084	0.197	0.089
$\tau = 0.50$								
$N = 1{,}000$	0.607	0.298	0.071	0.810	0.408	0.095	1.587	0.785
$N = 10{,}000$	0.444	0.241	0.045	0.695	0.393	0.077	1.338	0.769
$\tau = 0.75$								
$N = 1{,}000$	1.393	0.718	0.071	1.918	1.087	0.089	4.573	2.247
$N = 10{,}000$	0.982	0.571	0.045	1.705	1.038	0.071	3.862	2.254

Note. Root mean squared error (RMSE) of the posterior means across 100 samples for a Dirichlet process mixture of MSQR-CF (DPM-MSQR-CF), Bayesian multi-state quantile regression with cured fraction (MSQR-CF), and multi-state quantile regression (MSQR); at different quantiles τ and sample sizes N.

Table 4. Simulation study

	DPM-MSQR-CF	MSQR-CF	MSQR
$\tau = 0.25$			
$N = 1{,}000$	0.067168	0.065548	1.383605
$N = 10{,}000$	0.059675	0.062344	1.354225
$\tau = 0.50$			
$N = 1{,}000$	0.113985	0.119150	1.315509
$N = 10{,}000$	0.108583	0.113383	1.289967
$\tau = 0.75$			
$N = 1{,}000$	0.222629	0.234933	1.173807
$N = 10{,}000$	0.218935	0.229457	1.155595

Note. Relative root mean squared error (RRMSE) of estimated conditional quantiles across 100 samples for a Dirichlet process mixture of MSQR-CF (DPM-MSQR-CF), Bayesian multi-state quantile regression with cured fraction (MSQR-CF), and multi-state quantile regression (MSQR); at different quantiles τ and sample sizes N.

Even in a default prior setting the DPM-MSQR-CF model seems to be able to capture unobserved heterogeneity successfully, and avoid bias in the regression coefficients not modulated by the mixture. This is in contrast to other models, which all lead to higher RMSE values in all cases. Therefore, it is worth noting that the proposed model consistently outperforms the benchmark models across all configurations. The relatively higher RMSE values observed in some cases are primarily due to the limited information available in small-sample settings. As the sample size increases, the estimates become significantly more accurate and stable. When examining the relative root mean squared error of the predicted quantiles, see Table 4, the proposed models show excellent predictive performance even with $N = 1{,}000$, which further improves with larger sample sizes. The model by Farcomeni and Geraci (2020), on the other hand, fails as it does not account for the cured fraction, tending to overestimate the coefficients and produce inaccurate predictions.

4.3 An evaluation of the variance correction

We evaluated the empirical coverage of the credible intervals for parameters β_1 and β_2, which pertain to the AL component of the likelihood, and for γ_1, the parameter of the cure rate model, using

Table 5. Empirical 0.95 coverage rates for γ_1, β_1, and β_2 across 100 samples in Dirichlet process mixture of MSQR-CF (DPM-MSQR-CF)

	τ	β_1	β_2	γ_1
MCMC estimates				
$N = 1{,}000$	0.25	0.99	0.96	1.00
	0.50	0.98	0.65	1.00
	0.75	0.98	0.81	1.00
$N = 10{,}000$	0.25	0.95	0.84	1.00
	0.50	0.97	0.33	1.00
	0.75	0.97	0.68	1.00
Yang et al. (2016) estimates				
$N = 1{,}000$	0.25	1.00	1.00	1.00
	0.50	1.00	1.00	1.00
	0.75	1.00	1.00	1.00
$N = 10{,}000$	0.25	1.00	1.00	1.00
	0.50	1.00	1.00	1.00
	0.75	1.00	1.00	1.00

the result of the simulation study. We compute credible intervals for both the naive unadjusted estimator and the corrected one. The results are in Table 5. Uncorrected estimates have adequate 0.95 coverage rate when $N = 1{,}000$, but a marked deterioration is observed at $N = 10{,}000$. In particular, γ_1 consistently exhibits perfect coverage regardless of the sample size. This highlights that as N increases, the credible intervals for parameters linked to the AL component become increasingly less reliable from a frequentist perspective. On the other hand, the Yang et al. (2016) corrected estimator has a perfect coverage, regardless of the sample size.

However, it is important to highlight here that empirical coverage of credibility intervals depends on the prior choice. Bayesian credible intervals do not necessarily have frequentist properties, unless specific choices (e.g. matching priors) are made. In this light, it is not the actual values in Table 5 that matter, but the fact that with the correction the coverage substantially improves due to the validity of the corrected covariance estimates.

A possibly better evidence that Yang et al. (2016) correction is effective in our context can be obtained by comparing the distance (for instance, in terms of Frobenius norm) between the estimated and true asymptotic covariance matrices. Bernstein von Mises theorem (BvM) allows us to approximate the true posterior covariance as $(1/N)I(\theta_0)^{-1}$, where $I(\theta_0)$ is the observed information matrix evaluated at the true parameter value θ_0. We then compute the Frobenius distance between the true approximate asymptotic covariance matrix and the estimate covariance matrices, with and without correction. Table 6 shows that the correction proposed by Yang et al. (2016) consistently yields estimates that are closer to the true covariance matrix at each quantile level and sample size.

4.3.1 Partial batch MCMC

In order to numerically test the effectiveness of the proposed method for speeding up estimation with big data, we show a brief comparative study of standard and batched methods. We generated 100 samples of $N = 50{,}000$ observations in the same setting as the previous subsection; implementing both the standard MCMC estimator for DPM-MSQR-CF (i.e. when $\Omega = 1$) and the partial batch MCMC, with $\Omega = 5$ and $\Omega = 10$ blocks.

Table 7 shows that even with this sample size, the posterior mean and (corrected) posterior standard deviation are correctly approximated. The posterior mean is approximated better than the standard deviation, as expected. The approximation is better with $\Omega = 5$ than with $\Omega = 10$,

Table 6. Frobenius distance between the adjusted and unadjusted estimated covariance matrix of parameter estimates the true one

N	τ	Yang et al. (2016)	MCMC
1,000	0.25	0.002	0.092
	0.50	0.050	0.418
	0.75	0.688	2.204
10,000	0.25	0.001	0.035
	0.50	0.035	0.272
	0.75	0.586	1.735

Note. Results are the median over 100 replicates.

Table 7. Simulation study

		β_1	β_2	γ_1
$\tau = 0.25$				
$\Omega = 5$	M_1	1.057	1.004	1.015
$\Omega = 5$	M_2	0.944	0.924	1.081
$\Omega = 10$	M_1	1.004	1.012	1.040
$\Omega = 10$	M_2	0.918	0.863	1.133
$\tau = 0.50$				
$\Omega = 5$	M_1	0.985	0.976	1.022
$\Omega = 5$	M_2	0.883	0.897	0.960
$\Omega = 10$	M_1	0.985	0.959	1.017
$\Omega = 10$	M_2	0.765	0.881	0.961
$\tau = 0.75$				
$\Omega = 5$	M_1	0.994	0.937	1.002
$\Omega = 5$	M_2	0.923	0.846	0.856
$\Omega = 10$	M_1	0.944	0.875	1.016
$\Omega = 10$	M_2	0.844	0.739	0.822

Note. Geometric mean over 100 samples of the ratios between the posterior moments for standard MCMC and the (corrected) posterior moments for partial batch MCMC. M_1 indicates the posterior mean, M_2 the posterior standard deviation. The sample size is fixed as $N = 50{,}000$.

as could also be expected, due to the fact that smaller batches are used when $\Omega = 10$. Indeed, with a batch of only 5,000 observations, posterior consistency might not be close; which is why some ratios are not very close to the unity. If anything, the batched method seems to slightly overestimate variability, which is better than the reverse.

We speculate, also in light of further comparisons that we have done with simplified models and larger sample sizes, that with real big data the ratios would be much closer to the unity. Note that, in our application, the sample size is more than ten times the one of the samples simulated in this section.

Finally, it is worth highlighting the substantial computational cost advantage achieved through the implementation of partial batching. On a single dataset, we observed that in a standard computational setup, 100 iterations required about 125 s for the classical MCMC estimator for DPM-MSQR-CF, 65 s for the partial MCMC with $\Omega = 5$ and 11 s for the partial batch MCMC with $\Omega = 10$.

Table 8. Jail data

Code	Variable	Statistic
X_1	Gender	77.5% males
X_2	Age at first detention	35.62 ± 11.53
X_3	Violent crimes	66%
X_4	Sexual crimes	5.7%
X_5	Property Crimes	26.1%
	Drug Crimes	11.1%
X_6	African American	29.6%
X_7	Asian/Pacific	1.7%
X_8	Indigenous	0.9%
X_9	Other POC	11.1%
	White	56.7%
X_{10}	No. of arrests post 2019	1.18 ± 0.37
X_{11}	Latest detention duration	0.03 ± 1.95

Note. Covariates used in the analysis, with their code when included in the model, and summary statistics. For X_{10} and X_{11} we report the geometric mean $\pm$ the standard deviation for the log variable. 'Other POC' indicates 'other people of colour' beyond African Americans.

5 Analysis of the jail dataset

We implemented the proposed model on a sample of $N = 550{,}994$ individuals who were in jail at least once across the U.S. in the period 2020–2023, or were already in jail at the beginning of follow-up; for a total of $\sum n_i = 686{,}319$ observations. For each spell, we do not have information on the expected duration (i.e. the judgment). We model the transition from freedom to jail, with time expressed in years. For individuals who were not in jail at the beginning of follow-up, we discard the first observed freedom spell so that all subjects in our sample have at least one documented detention. About 85% of the individuals are censored for their last freedom spell (that is, they are outside of jail at the end of follow-up). More than 16% of the individuals have more than one detention spell during follow-up; of these, about 20% have three spells, 5% have four, and 3% have five or more. Restricting to the events, freedom spells are usually rather short; with a median of about 6 months, third quartile of about 1 year, and last decile of about 19 months.

The covariates that will be used to model both the conditional time to transition and the probability of cure are reported in Table 8; together with some descriptive statistics. These are in line with the descriptives reported in Western et al. (2021), with some interesting differences. In fact, data from Western et al. (2021) are restricted to New York City and the years 2008 to 2017. Our nation-wide and slightly more recent data report an about 10% smaller proportion of males, and a much larger proportion of whites. This is in line with a trend of gender gap reduction for jail detention over time, and a huge spatial heterogeneity for race. In fact, if we restrict our data to New York, we obtain about 60% African-American, which is in line with Western et al. (2021). The covariates of the crime category refer to the reason for the previous arrest; more than one reason is possible, even if a unique category is recorded for more than 91% of arrests, 8% have two reasons, and only 0.5% have three or four.

We model the logarithm of the transition times to ensure that the linear combination of the covariates falls within the support of the variable; and to make the assumption that the centring measure G_0 is a bivariate Gaussian more credible. Transition times (conditionally on $R_i^{(mj)} = 0$) are modelled at quantiles $\tau = 0.25, 0.5, 0.75$.

For ease of interpretation, we further constrain our proposed model in two respects. First, we select a common number of latent states at different quantiles. To do so, we used the DPM sampling at $\tau = 0.5$. A more formal alternative would have been to combine the log-likelihoods at all

Table 9. Jail data

	WAIC
DPM-MSQR	−7,232.934
DPM-MSQR-CF	**−7,325.079**

Note. WAIC calculated on the Dirichlet Process Mixture multi-state quantile regression model with and without a cured fraction.

Table 10. Jail data

Time to transition											
	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}
$\tau = 0.25$											
$E(\cdot \mid X)$	−0.062	0.003	−0.092	0.571	−0.452	−0.078	−0.473	−0.026	0.030	−0.384	−0.117
$SD(\cdot \mid X)$	0.023	0.028	0.086	0.105	0.129	0.007	0.126	0.106	0.015	0.133	0.108
$\tau = 0.50$											
$E(\cdot \mid X)$	−0.029	−0.000	0.039	0.170	−0.398	0.352	0.130	−0.010	−0.526	−0.161	−0.100
$SD(\cdot \mid X)$	0.069	0.028	0.015	0.041	0.035	0.053	0.020	0.120	0.080	0.055	0.053
$\tau = 0.75$											
$E(\cdot \mid X)$	0.024	−0.024	−0.452	0.604	−0.345	0.363	−0.182	0.358	−0.404	0.250	0.140
$SD(\cdot \mid X)$	0.331	0.040	0.274	0.224	0.278	0.299	0.041	0.055	0.082	0.318	0.024
Cure model	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}	γ_{11}
$E(\cdot \mid Z)$	−0.217	0.009	−0.087	−0.011	−0.173	0.052	0.181	−0.145	−0.331	−0.188	−0.030
$SD(\cdot \mid X)$	1.891	0.207	2.389	2.363	2.195	1.817	0.937	1.146	2.166	2.337	0.827

Note. Posterior means and standard deviations for the DPM-MSQR-CF at $\tau = 0.25$, $\tau = 0.50$, $\tau = 0.75$.

quantiles, but we have seen that it would likely not have led to different results, and it would have involved larger computational times and, therefore, smaller number of iterations for the sampler. Secondly, we constrain the latent indicators and parameters for the cure rate model to be equal at different quantiles. We thus have a common probability of occurrence for recidivism, regardless of the quantile of interest for transition times; and each subject belongs to the same latent state at different quantiles (clearly, state-specific parameters are still quantile dependent). This is useful for interpretation and not very restrictive. In the unconstrained case, as could be expected, the posterior means of these parameters were actually close to each other.

We have fit models with (DPM-MSQR-CF) and without (DPM-MSQR) a cure fraction. Posterior distributions are estimated using batch MCMC, with $\Omega = 50$ blocks. The prior setting for both quantile regression and cure rate model parameters follows the approach introduced in Section 3. We compare DPM-MSQR-CF with DPM-MSQR in terms of the Widely Applicable Information Criterion (WAIC, Watanabe, 2013), in the definition of Gelman et al. (2014); reported in Table 9. It can be seen that the inclusion of a cure fraction provides a substantially better fit. This further motivates our contribution; and it is a rather common phenomenon, in our experience, when the event rate is relatively low. As already stated, we should be wary of interpreting 'cured' individuals as 'rehabilitated'. It is indeed true that the model provides an estimate of the proportion of individuals that might not experience a new detention, but this proportion includes both individuals who will not commit other crimes (or, at least, not be suspected of any) and individuals who will enter a prison (possibly even more than once).

Posterior summaries for the manifest distribution of our proposed model are reported in Table 10, and in Table 11 we show the results related to the latent intercepts and weights.

Table 11. Jail data

Time to transition			
$\tau = 0.25$	$k = 1$	$k = 2$	$k = 3$
$E(\beta_0 \mid X)$	0.245	−0.668	−1.617
$\mathrm{SD}(\beta_0 \mid X)$	0.023	0.022	0.028
$\tau = 0.50$			
$E(\beta_0 \mid X)$	0.741	−0.375	−1.371
$\mathrm{SD}(\beta_0 \mid X)$	0.011	0.013	0.020
$\tau = 0.75$			
$E(\beta_0 \mid X)$	1.514	0.729	−0.188
$\mathrm{SD}(\beta_0 \mid X)$	0.018	0.018	0.026
Cure model			
$E(\gamma_0 \mid X)$	1.471	−1.234	0.393
$\mathrm{SD}(\gamma_0 \mid X)$	0.009	0.058	0.090
Mixture weights			
$E(w \mid X)$	0.797	0.130	0.072
$\mathrm{SD}(w \mid X)$	0.012	0.013	0.014

Note. Posterior means and standard deviations for the intercepts and mixture weights of the DPM-MSQR-CF model, at $\tau = 0.25$, $\tau = 0.50$, $\tau = 0.75$.

There is clear evidence of unobserved heterogeneity, as could be expected, with the model estimating three well-separated latent classes.

The majority of individuals are assigned to the first group (79.7%), with baseline non-cured individuals returning to jail in a median time of exp (0.741) ≈ 2.1 years, within an interval of approximately 1.28 years to 4.54 years (first and third quartile, respectively). In contrast, 13% of individuals belong to the second group, with recurrence times for baseline individuals ranging from 6 months to 2 years (first and third quartile, respectively) and a median time to recurrence of approximately 8.5 msonths. Lastly, the third group, consisting of 7.2% of individuals, has the shortest recurrence time, ranging from 2.4 to 10 months, with a median time to recurrence of 3 months.

Men appear to have slightly shorter recurrence times in the first quartile (6%) and median (3%) than females. Age at first detention appears to have almost no association with the time outcome. There is some association of recurrence times with the type of crime and race. Compared to baseline (drug crimes), recidivist sexual offenders (β_4) tend to stay free for the longest time in all quantiles. This result could be associated with policies for sex offenders, which have received considerable attention, especially in recent years (Schmucker & Lösel, 2017). Focusing on β_5, it can be seen that inmates charged with property crimes, controlling for other variables, have shorter recurrence times in all quantiles.

The association of recurrence time with ethnicity is more complex. There does not seem to be much difference among races at $\tau = 0.25$, i.e. race does not matter much for people at high risk of short-term recidivism; with possibly just the exception of Asian Americans (β_7), who have slightly shorter recurrence times. At the median and third quartile, African Americans (β_5) have longer recurrence times; while other POC (β_9) have shorter recurrence times. Finally, indigenous individuals (β_8) appear to have differential (slightly longer) recurrence times compared to white individuals only in the third quartile.

Our last, possibly the main, result of interest involves the expected number of arrests over time, reported in Figure 1. The results are stratified by latent class and certain values of the covariates for comparison.

Cumulative hazard functions are very well separated when stratifying by latent class. Unobserved heterogeneity clearly plays the most important role, with only a minor difference in cumulative hazard brought about by the covariates. While for one group, to which about 80%

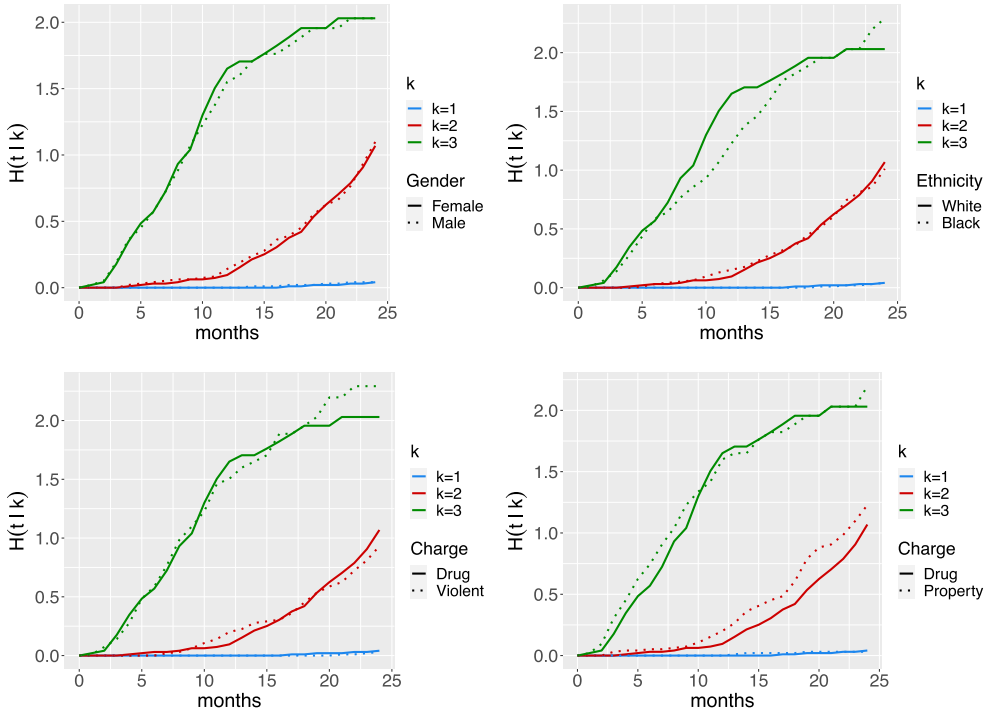

Figure 1. Jail data. Cumulative hazard function conditional to observed clusters and certain covariates.

of the subjects are assigned, jail recurrence is unlikely within the first 2 years; the other two groups accumulate more than one detention per year of freedom. One group has an initial high risk, with about 1.75 detentions in the first year of freedom, which then decreases in magnitude to reach about 2 detentions within the first 2 years of freedom. Another latent class is associated with a low risk in the first year (about 0.25 detentions in the first year of freedom), which then increases to reach more than one new detention within the first 2 years of freedom.

As an additional benchmark, Table 12 reports the estimated coefficients from the multi-state quantile regression model without cured fraction proposed by Farcomeni and Geraci (2020). Since the WAIC is a Bayesian criterion that requires a fully probabilistic specification, it cannot be computed for this frequentist model. Nevertheless, the substantial discrepancies observed in the estimated coefficients further highlight the capacity of our approach to account for unobserved heterogeneity, thereby enabling more robust inference. For instance, consider inmates charged with sexual offenses: in our model, the estimated coefficient β_4 is consistently positive across all quantiles, whereas in this case it is markedly smaller and nearly null at the median. Similarly, for inmates charged with property crimes, our model estimates a negative value for all the considered τ levels, while in the benchmark model a positive coefficient is found at the upper quartile ($\hat{\beta}_5 = 0.149$), suggesting potential misclassification when latent heterogeneity is not explicitly modelled.

6 Conclusions

We proposed a Dirichlet process mixture of multi-state quantile regression models incorporating a cure fraction. Our model is a quantile counterpart to Conlon et al. (2014); with the addition of unobserved heterogeneity. It also extends directly Farcomeni and Geraci (2020) in several directions. As briefly mentioned above, there are advantages in using a quantile regression approach to multi-state modelling, prominently interpretability and computational convenience. We have specified a latent class model in which individuals do not change the assigned class over time. Dynamic latent variable models would be much more cumbersome, and are left for further work (see e.g. Barone et al., 2024 for an example for recurrent event models).

Table 12. Estimates and standard errors for the parameters of the MSQR model of Farcomeni and Geraci (2020) at $\tau = 0.25$, 0.50, and 0.75

Time to transition	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}
$\tau = 0.25$												
Estimate	0.038	−0.105	0.002	−0.064	0.149	−0.163	0.095	−0.036	−0.256	−0.006	−0.311	0.083
Standard error	0.102	0.022	0.001	0.033	0.051	0.052	0.029	0.104	0.089	0.028	0.084	0.022
$\tau = 0.50$												
Estimate	0.729	−0.040	0.002	0.016	0.013	−0.109	−0.007	−0.164	0.078	−0.102	−0.254	−0.113
Standard error	0.067	0.025	0.001	0.054	0.066	0.066	0.017	0.092	0.193	0.025	0.066	0.022
$\tau = 0.75$												
Estimate	0.877	0.027	0.000	0.225	0.159	0.149	−0.038	−0.405	−0.617	−0.224	−0.195	−0.195
Standard error	0.108	0.076	0.001	0.083	0.108	0.074	0.050	0.282	0.294	0.094	0.024	0.088

Our multi-state framework encompasses several scenarios, including recurrent event, competing, and semi-competing risk data.

A Bayesian approach is used, accounting for unobserved heterogeneity without specifying the number of latent classes in advance. A simple batching strategy is used to scale the MCMC procedure for big data, as we work with more than half a million individuals. Even after batching, approximately nine hours are needed to obtain a sample from the posterior for a fixed quantile. Using 40 CPUs in parallel, model fitting took approximately 1 day for obtaining samples at the pre-specified grid of quantiles. One the other hand, classical methodologies for multi-state modelling, both with and without a cure fraction, would be computationally intractable for our application.

Similarly, frequentist inference for the proposed model might be feasible, but a little more computationally intensive and less stable than the Bayesian one we have adopted. A clear advantage of the Bayesian approach, beyond the possibility to incorporate prior knowledge, is the fact that we can effectively specify an infinite mixture model; without having to select the number of latent classes.

We have shown that, as often happens with complex time-to-event data with low event rates, a cure-rate model is advantageous over one that assumes that all individuals will eventually experience the event. See also Jackson et al. (2022) on this point. In our motivating application, as sometimes happens, the cure fraction is mostly used to improve the goodness of fit. It would be cumbersome indeed to evaluate the role of cured individuals: individuals that are not expected to reenter jail might indeed enter a prison, an information that is not present in our data. However, there are several other application domains (e.g. breast cancer) with direct interpretability of the cure fraction.

Our model was applied to the study of transitions from freedom to jail in the period 2020–2023 in the U.S., on a sample of individuals who have previously experienced detention; thus updating the literature on jail recidivism in the U.S. In addition, to our knowledge, many studies on jail recidivism are limited to local (e.g. state) data.

The estimated cumulative hazard is one of the key by-products of our procedure and should be interpreted with caution. In fact, it does not take into account the distribution of any other transition, including the one *back* to the entry state. This is common to any cumulative hazard estimate in the multi-state context, but it is worth underlining. In this sense, an expected number of two events per year shall be interpreted as two events per year *of freedom*, or per year plus jail or prison time for each detention. We believe that not counting the time spent in jail before any further detention makes the figure more clearly interpretable. Anyway, estimates for the cumulative number of events could readily be adjusted.

We have found a relatively low cumulative hazard for 80% of the population under study, but a high risk for the remaining 20%; with short-term recurrence and up to two detentions per year of

freedom. Time-to-transition between freedom and a new jail spell was estimated as very low at the first quartile. Little association was found with the extant covariates. It should be mentioned that latent clustering does not seem also to be associated, and hence possibly explained, by the available covariates (results not shown). Notably, the covariates are on the other hand very heterogeneous across space, especially race and type of crime. The main conclusion is that for about one fifth of the population experiencing detention at least once, jail recurrence is a persistent event, with very short free time between spells. A characterization of this population can not be easily achieved with the extant covariates. If recidivism is used to measure the correctional effectiveness of prisons and jails, the evidence suggests that jails—and the programs and initiatives targeting individuals with jail experience—perform even worse than prisons.

Acknowledgments

The authors thank the NYU Public Safety Lab for sharing the Jail Data Initiative data. The authors also thank two reviewers and an associate editor for constructive remarks.

Conflicts of interest: The authors declare no conflict of interest.

Funding

The first author was supported by the European Union-NextGenerationEU, in the framework of the GRINS-Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 - CUP E83C22004690001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them. The second author was supported in part by the PRIN funding scheme of the Italian Ministry of University and Research (Grant No. 2022LANNKC CUP E53D23005810006).

Data availability

The data underlying this article were provided by the NYU Public Safety Lab's Jail Data Initiative (https://jaildatainitiative.org). The authors do not have permission to share the data. Access to the data may be requested directly from the Jail Data Initiative.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series A*.

References

Barone R., Farcomeni A., & Mezzetti M. (2024). Latent Markov time-interaction processes. *Journal of Computational and Graphical Statistics*, *34*(3), 984–993. https://doi.org/10.1080/10618600.2024.2421984

Beaudry G., Yu R., Perry A. E., & Fazel S. (2021). Effectiveness of psychological interventions in prison to reduce recidivism: A systematic review and meta-analysis of randomised controlled trials. *The Lancet: Psychiatry*, *8*(9), 759–773. https://doi.org/10.1016/S2215-0366(21)00170-X

Beyersmann J., & Schumacher M. (2008). A note on nonparametric quantile inference for competing risks and more complex multistate models. *Biometrika*, *95*(4), 1006–1008. https://doi.org/10.1093/biomet/asn044

Bottai M., Cai B., & McKeown R. E. (2010). Logistic quantile regression for bounded outcomes. *Statistics in Medicine*, *29*(2), 309–317. https://doi.org/10.1002/sim.v29:2

Conlon A. S. C., Taylor J. M. G., & Sargent D. J. (2014). Multi-state models for colon cancer recurrence and death with a cured fraction. *Statistics in Medicine*, *33*(10), 1750–1766. https://doi.org/10.1002/sim.v33.10

Durose M. R., Cooper A. D., & Snyder H. N. (2014). *Recidivism of prisoners released in 30 states in 2005: Patterns from 2005 to 2010*. Vol. 28. US Department of Justice, Office of Justice Programs, Bureau of Justice.

Escobar M. D., & West M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, *90*(430), 577–588. https://doi.org/10.1080/01621459.1995.10476550

Farcomeni A., & Geraci M. (2020). Multistate quantile regression models. *Statistics in Medicine*, *39*(1), 45–56. https://doi.org/10.1002/sim.v39.1

Gelman A., Hwang J., & Vehtari A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. https://doi.org/10.1007/s11222-013-9416-2

Geraci M., & Jones M. C. (2015). Improved transformation-based quantile regression. *Canadian Journal of Statistics*, *43*(1), 118–132. https://doi.org/10.1002/cjs.v43.1

Golder S., Ivanoff A., Cloud R. N., Besel K. L., McKiernan P., Bratt E., & Bledsoe L. K. (2005). Evidence-based practice with adults in jails and prisons. *Best Practices in Mental Health*, *1*(2), 100–132. https://doi.org/10.70256/102376jxqetp

Hsieh J.-J., & Wang H.-R. (2018). Quantile regression based on counting process approach under semi-competing risks data. *Annals of the Institute of Statistical Mathematics*, *70*(2), 395–419. https://doi.org/10.1007/s10463-016-0593-6

Jackson C. H., Tom Brian D. M., Kirwan P. D., Mandal S., Seaman S. R., Kunzmann K., Presanis A. M., & Angelis D. D. (2022). A comparison of two frameworks for multi-state modelling, applied to outcomes after hospital admissions with COVID-19. *Statistical Methods in Medical Research*, *31*(9), 1656–1674. https://doi.org/10.1177/09622802221106720

Kalli M., Griffin J. E., & Walker S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, *21*(1), 93–105. https://doi.org/10.1007/s11222-009-9150-y

Koenker R., & Bassett G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, *46*(1), 33–50. https://doi.org/10.2307/1913643

Lange S., Rehm J., & Popova S. (2011). The effectiveness of criminal justice diversion initiatives in North America: A systematic literature review. *International Journal of Forensic Mental Health*, *10*(3), 200–214. https://doi.org/10.1080/14999013.2011.598218

Lavigne A., & Liverani S. (2024). Quantifying the uncertainty of partitions for infinite mixture models. *Statistics & Probability Letters*, *204*(2), 109930. https://doi.org/10.1016/j.spl.2023.109930

Li R., & Peng L. (2011). Quantile regression for left-truncated semicompeting risks data. *Biometrics*, *67*(3), 701–710. https://doi.org/10.1111/biom.2011.67.issue-3

MacEachern S. N., & Müller P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics*, *7*(2), 223–238. https://doi.org/10.1080/10618600.1998.10474772

Mears D. P., Cochran J. C., Bales W. D., & Bhati A. S. (2016). Recidivism and time served in prison. *The Journal of Criminal Law and Criminology*, *106*, 83–124. https://scholarlycommons.law.northwestern.edu/jclc/vol106/iss1/5

Meira-Machado L., de Uña Alvarez J., Cadarso-Suarez C., & Andersen P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research*, *18*(2), 195–222. https://doi.org/10.1177/0962280208092301

Mu Y. M., & He X. M. (2007). Power transformation toward a linear regression quantile. *Journal of the American Statistical Association*, *102*(477), 269–279. https://doi.org/10.1198/016214506000001095

Peng L., & Fine J. P. (2009). Competing risks quantile regression. *Journal of the American Statistical Association*, *104*(488), 1440–1453. https://doi.org/10.1198/jasa.2009.tm08228

Peng Y., & Yu B. (2021). *Cure models: Methods, applications, and implementation*. Chapman and Hall/CRC.

Qu L., Sun L., & Sun Y. (2023). A mark-specific quantile regression model. *Biometrika*, *111*(1), 255–272. https://doi.org/10.1093/biomet/asad039

Rosenthal J. S. (2011). Optimal proposal distributions and adaptive MCMC. In *Handbook of Markov Chain Monte Carlo*, *4*(10.1201).

Sawyer W., & Wagner P. (2022). *Mass incarceration: The whole pie* (Technical Report). Prison Policy Initiative.

Schmucker M., & Lösel F. (2017). Sexual offender treatment for reducing recidivism among convicted sex offenders: A systematic review and meta-analysis. *Campbell Systematic Reviews*, *13*(1), 1–75. https://doi.org/10.4073/csr.2017.8

Sun X., Peng L., Huang Y., & Lai H. J. (2016). Generalizing quantile regression for counting processes with applications to recurrent events. *Journal of the American Statistical Association*, *111*(513), 145–156. https://doi.org/10.1080/01621459.2014.995795

Wade S., & Ghahramani Z. (2018). Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Analysis*, *13*(2), 559–626. https://doi.org/10.1214/17-BA1073

Walker S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics: Simulation and Computation*, *36*(1), 45–54. https://doi.org/10.1080/03610910601096262

Watanabe S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research*, *14*, 867–897. https://doi.org/10.48550/arXiv.1208.6338

Western B., Davis J., Ganter F., & Smith N. (2021). The cumulative risk of jail incarceration. *Proceedings of the National Academy of Sciences*, *118*(16), e2023429118. https://doi.org/10.1073/pnas.2023429118

Wu T.-Y., Wang R. Y. X., & Wong W. H. (2022). Mini-batch Metropolis–Hastings with reversible SGLD proposal. *Journal of the American Statistical Association*, *117*(537), 386–394. https://doi.org/10.1080/01621459.2020.1782222

Yang Y., Wang H. J., & He X. (2016). Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. *International Statistical Review*, *84*(3), 327–344. https://doi.org/10.1111/insr.v84.3

Yu K., & Moyeed R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, *54*(4), 437–447. https://doi.org/10.1016/S0167-7152(01)00124-9