ECOM168: Bayesian generalised linear models

Daniele Bianchi

School of Economics and Finance,
Queen Mary University of London

Outline

1. Introduction
2. Bayesian Probit regression
3. Bayesian Poisson regression
4. Bayesian quantile regression

Introduction

Introduction

Generalised linear models are extensions of the linear regression model described in the previous lectures.

- $\hookrightarrow$ key idea: turns covariates into a real number via linear projection and then transform this value to fit the support of the response.

We will focus on three types of generalised linear models:

- $\hookrightarrow$ Limited dependent variable (Probit/logit).
- $\hookrightarrow$ Count data (Poisson regression).
- $\hookrightarrow$ Quantile regressions.

On the methodological side we discuss the Metropolis-Hastings algorithm, which complements the Gibbs sampler for the simulation of complex distributions.

Motivation

So far we modelled the connection between a response variable y and a set of predictors X by a linear dependence relation of the form

$$y = X\beta + \varepsilon, \qquad \varepsilon \sim N\left(0, \sigma^2 \mathbf{I}_n\right), \tag{1}$$

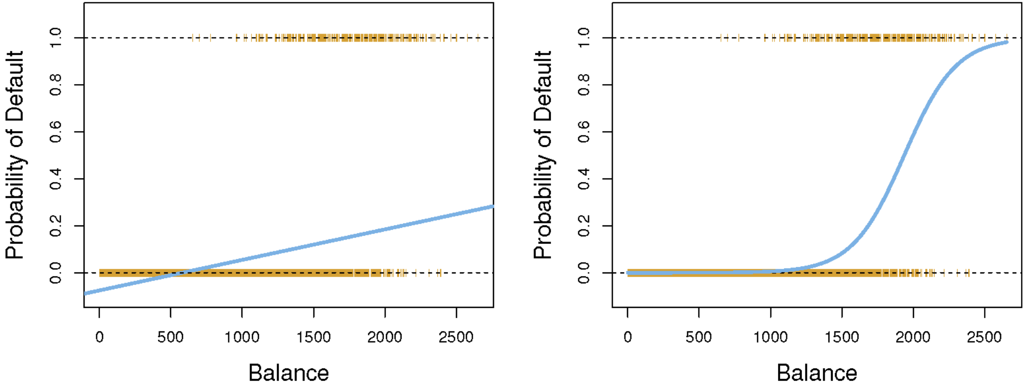
where $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ is an $n \times p$ matrix of covariates and $\beta = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ is the p-dimensional vector of regression parameters.

However, there are many instances where both the linearity and the normality assumptions are not appropriate, especially when the support of $y \in \mathbb{R}_+$ is restricted.

$\hookrightarrow$ for instance, $y = \{0, 1\}$ as it represents an indicator of occurrence of a particular event.

When the support of the dependent variable is restricted, a linear conditional expectation $\mathbb{E}\left[y|X, \beta\right] = X\beta$ could be fairly cumbersome to obtain.

Introduction



The orange marks indicate the response y, either 0 or 1. Linear regression (left panel) does not estimate $Pr\,(y = 1|X)$ well. Logistic regression (right panel) seems more suited for the task.

Introduction

We need a broader class of models to cover various dependence structures: that is, *generalised linear models* (GLM).

GLM stems from the fact that the dependence between y and X is partly *linear*; the conditional distribution of $y|X$ is defined in terms of $X\beta$,

$$y|\beta \sim p\left(y|X\beta\right),$$

In general terms, a GLM is specified by two functions:

- $\hookrightarrow$ A conditional density p of y given X that belongs to an exponential family.
- $\hookrightarrow$ A *link* function g that relates the mean $\mu = \mathbb{E}\left[y|X\right]$ and the covariate vector X, i.e.,

$$\mathbb{E}\left[y|\beta, \sigma^2\right] = g^{-1}\left(X\beta\right),$$

Introduction

The ordinary linear regression is obviously a special case of GLM where $g(X) = X$ and $y|\beta, \sigma^2 \sim N(X\beta, \sigma^2)$.

However, outside the linear model, the interpretation of the coefficients β is much more delicate because these coefficients do not relate directly to the observables;

$\hookrightarrow$ due to the presence of a link function that cannot be the identity.

For instance, in the logistic regression model (defined in the following section), the linear dependence is defined in terms of the *log-odds ratio* $\log(\pi/(1-\pi))$.

Bayesian Probit regression

Probit model

For binary response variables, a possible link function is the *probit transform*, $g(\mu) = \Phi^{-1}(\mu)$, where Φ is the standard normal cumulative distribution function.

The corresponding likelihood is given by

$$p(y|X,\beta) \propto \prod_{i=1}^{n} \Phi\left(x_i^\top \beta\right)^{y_i} \left[1 - \Phi\left(x_i^\top \beta\right)\right]^{1-y_i},$$

The key advantage of the Probit model is computational tractability.

$\hookrightarrow$ Observing $y_i = 1$ corresponds to the case $z_i \geq 0$, where z_i is a latent (meaning unobserved) variable such that $z_i \sim N\left(x_i^\top \beta, 1\right)$.

$\hookrightarrow$ That is $y = \mathbb{I}(z_i \geq 0)$ appears as a dichotomized linear regression response.

Probit model

If no prior information is available we can consider a flat prior $p(\beta) \propto 1$,

$\hookrightarrow$ in this case the posterior distribution is equal to the likelihood, i.e., $p(\beta|y) = p(y|\beta)$.

However, under a flat prior the conditional posterior cannot be sampled from directly, and must be simulated using a so-called Metropolis-Hastings (MH) algorithm.

A variety of MH algorithms have been proposed; in the following we consider a sampler that appears to work reasonably well in small-dimensional cases.

Probit MH sampler

Initialization: Compute the MLE $\widehat{\beta}$ and the covariance matrix $\widehat{\Sigma}$ corresponding to the asymptotic covariance of $\widehat{\beta}$, and set $\beta^{(0)} = \widehat{\beta}$.

Iteration $t \geq 1$:

1. Generate $\beta^* \sim N\left(\beta^{(t-1)}, \tau^2\widehat{\Sigma}\right)$
2. Compute

$$\alpha\left(\beta^{(t-1)}, \beta^*\right) = \min\left(1, \frac{p\left(\beta^*|y\right)}{p\left(\beta^{(t-1)}|y\right)}\right),$$

3. With probability $\alpha\left(\beta^{(t-1)}, \beta^*\right)$, take $\beta^{(t)} = \beta^*$, otherwise $\beta^{(t)} = \beta^{(t-1)}$.

Probit with Gibbs

While the MH algorithm is numerically convenient, it is quite problematic in high dimensions.

A popular alternative builds upon a linear regression representation of the Probit based on a latent variable $z = (z_1, \ldots, z_n)^\top$ such that

$$z \sim N\left(X\beta, \sigma^2\right), \qquad \text{with} \qquad y = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases} \tag{2}$$

Probit with Gibbs

Assuming a normal prior $\beta \sim N(0, D)$, Albert and Chib (1993) showed that the conditional distribution of $\beta|z$ can be defined as $\beta|z \sim N(\mu_\beta, \Sigma_\beta)$, with

$$\Sigma_\beta = \left(D^{-1} + X^\top X\right)^{-1}, \qquad \mu_\beta = \Sigma_\beta X^\top z,$$

while the conditional distribution of z given β is a truncated Normal,

$$z|\beta \propto \begin{cases} z_i \sim TN_{(-\infty,0)}\left(x_i^\top \beta\right) & \text{if } y_i = 0 \\ z_i \sim TN_{(0,\infty)}\left(x_i^\top \beta\right) & \text{if } y_i = 1 \end{cases} \tag{3}$$

Assuming D is unknown a priori, the full conditional distribution takes the form

$$D|z, y, \beta \propto \prod_{j=1}^{p} IG\left(\frac{d_1 + 1}{2}, \frac{2}{d_2 + \beta_j^2}\right),$$

Bayesian Poisson regression

Bayesian Poisson regression

Let $y_1, \ldots, y_n$ be a sequence of count data, observed at discrete, evenly spaced time points.

A conventional model would be $y_t|\lambda_t \sim \mathsf{Po}\left(\lambda_t\right)$ where λ_t depends on a series of covariates $\boldsymbol{x}_t$;

$$y_t|\lambda_t \sim \mathsf{Po}\left(\lambda_t\right), \qquad \lambda_t \sim \exp\left(\boldsymbol{x}_t'\beta\right), \tag{4}$$

This is a conditional Poisson regression model. A conventional assumption is $\beta \sim N\left(0, D\right)$.

One may derive the conditional posterior density $p\left(\beta|y\right)$, but in general such distribution does not belong to a well-known distribution family.

Bayesian Poisson regression

Although $\log \lambda_t$ is linear in βs, the presence of a Poisson distribution in Eq.(4) causes non-normality as well as non-linearity of the mean λ_t in β.

Frühwirth-Schnatter and Wagner (2006) propose an interesting framework to address both non-normality and non-linearity in a Poisson regression by mean of two latent processes.

Key advantage: the introduction of these auxiliary, latent, processes allows to estimate a Poisson regression via a standard Gibbs sampler.

Data augmentation for Poisson regressions

Intuition 1: $y_t|\lambda_t$ can be regarded as the distribution of the number of jumps in the internal $[0, 1]$ of an unobserved Poisson process with intensity λ_t.

The first step of the data augmentation process in Frühwirth-Schnatter and Wagner (2006) creates such a Poisson process by introducing the inter-arrival times $\tau_{ij} \sim \text{Exp}\,(\lambda_t)$, such that

$$\tau_{ij}|\beta = \frac{\xi_{tj}}{\lambda_t}, \qquad \xi_{tj} \sim \text{Exp}(1),$$

This can be reformulated as a linear model

$$-\log \tau_{tj}|\beta = \boldsymbol{x}_t'\beta + \epsilon_{tj}, \tag{5}$$

where $\epsilon_{tj} = -\log \xi_{tj}$ with $\xi_{tj} \sim \text{Exp}\,(1)$.

Data augmentation for Poisson regressions

Intuition 2: The error term in Eq. may be regarded as the negative of the logarithm of an Exp(1) random variable, the density $p_\epsilon(\epsilon)$ of which is non-Gaussian.

Frühwirth-Schnatter and Wagner (2006) show that we can obtain a model that is conditionally Gaussian, we approximate this non-Normal density by a mixture of R normal components,

$$p_\epsilon(\epsilon) \approx \sum_{r=1}^{R} \omega_r N\left(\epsilon | m_r, s_r^2\right),$$

where m_r and s_r^2 are the mean and the variance of the Gaussian density $N\left(\epsilon | m_r, s_r^2\right)$.[1]

The approximate parameters $\left(\omega_r, m_r, s_r^2\right)$ are from Frühwirth-Schnatter and Wagner (2006).

[1] A similar approach to stochastic volatility has been adopted by Kim et al. (1998) and Chib et al. (2002).

Data augmentation for Poisson regressions

Given the mixture approximation, the second step of the data augmentation proposed by Frühwirth-Schnatter and Wagner (2006) introduces for each ϵ_{tj} the latent indicator r_{tj} as missing data.

Conditional on τ_{tj} and r_{tj}, the non-Normal, non-linear model in Eq.(4) reduces to a linear, Normal, regression of the form

$$-\log \tau_{tj}|\beta = \boldsymbol{x}_t'\beta + m_{r_{tj}} + \epsilon_{tj}, \qquad \epsilon_{tj}|r_{tj} \sim N\left(0, s^2_{r_{tj}}\right), \tag{6}$$

such that the posterior $p\left(\beta|\ldots\right)$ is proportional to a multivariate density.

$\hookrightarrow$ The key novelty of the Gibbs sampler is the sampling of the inter-arrival times $\tau = \{\tau_{tj}, j = 1, \ldots, y_t + 1\}$ and the component indicators $S = \{t_{tj}, j = 1, \ldots, y_t + 1\}$.

Data augmentation for Poisson regressions

While sampling $\beta|\ldots$ is model-dependent and relatively standard, the sample of $\tau, S|\ldots$ deserves some scrutiny.

The joint posterior $p\left(\tau, S|y, \beta\right)$ can be decomposed as

$$p\left(\tau, S|y, \beta\right) = p\left(S|\tau, y, \beta\right) p\left(\tau|y, \beta\right),$$

We discuss in turn how to sample these two components.

Data augmentation for Poisson regressions

Sampling inter-arrival times: Given β, y, the inter-arrival times are independent for different, such that, observations, ;

$$p\left(\tau|\beta, y\right) = \prod_{t=1}^{T} p\left(\tau_{t1}, \ldots, \tau_{t,y_t+1}|y_t, \beta\right),$$

For fixed t, the inter-arrival times $\tau_{t1}, \ldots, \tau_{t,n+1}$, where $n = y_t$ are stochastically dependent,

$$p\left(\tau_{t1}, \ldots, \tau_{t,n+1}|y_t = n, \beta\right) = p\left(\tau_{t,n+1}|y_t = n, \beta, \tau_{t1}, \ldots, \tau_{tn}\right) p\left(\tau_{t1}, \ldots, \tau_{tn}|y_t = n\right),$$

The joint distribution $p\left(\tau_{t1}, \ldots, \tau_{tn}|y_t = n\right)$ is approximated sampling the order statistics $u_t(1), \ldots, u_t(n) \sim$ Unif$(0, 1)$ of $n = y_t$, and define $\tau_{tj} = u_t(j) - u_t(j-1)$, for $j = 1, \ldots, n$ and $u_t(0) = 0$.

Conditional on y_t, only $n = y_t$ arrivals occur in $[0, 1]$, so that the arrival at $n + 1$ is known to occur after 1. As a result, $\tau_{t,n+1} = 1 - \sum_{j=1}^{n} \tau_{tj} + \xi_t$ with $\xi \sim$ Exp$\left(\lambda_t\right)$.

Data augmentation for Poisson regressions

Sampling indicators S: to sample the indicators S from $p\left(S|\tau, y, \beta\right)$, Frühwirth-Schnatter and Wagner (2006) use the fact that all indicators are conditionally independent given τ, y, β:

$$p\left(S|\tau, y, \beta\right) = \prod_{t=1}^{T} \prod_{j=1}^{y_t+1} p\left(r_{tj}|\tau_{tj}, \beta\right),$$

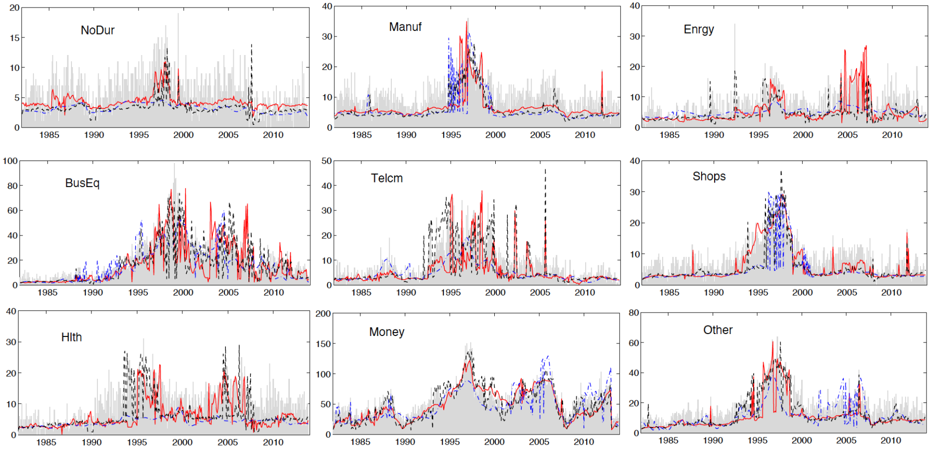
Thus, for each $t = 1, \ldots, T$ and $j = 1, \ldots, y_t + 1$ the indicator r_{tj} is sampled independently from

$$p\left(r_{tj} = k|\tau_{tj}, \beta\right) \propto p\left(\tau_{tj}|r_{tj} = k, \beta\right) \omega_k,$$

where

$$p\left(\tau_{tj}|r_{tj} = k, \beta\right) \propto \frac{1}{s_k} \exp\left\{ -\frac{1}{2}\left(\frac{-\log \tau_{tj} - \log \lambda_t - m_k}{s_k}\right)^2 \right\},$$

Example: Modeling merger waves



Industry merger activity and Poisson regression estimates of intensity rates. Source: Bianchi and Chiarella (2019).

Bayesian quantile regression

Bayesian quantile regression

Quantile regressions generalise traditional linear regression models by fitting a distinct set of parameters for each quantile of the distribution of the response variable.

The main distinction between the two models is:

- $\hookrightarrow$ Least squares allows to fit the conditional mean based on a set of parameters.
- $\hookrightarrow$ Quantile regression allows to fit the conditional distribution with one set of parameters for each quantiles.

Bayesian quantile regression

More formally, a regression specification can be represented as

$$y = f\left(y|X\right) + \varepsilon,$$

where $f\left(y|X\right)$ is a conditional mean function for y conditional on X.

Within the context of a linear regression model we have that $f\left(y|X\right) = \mathbb{E}\left(y|X\right) = X\beta$.

$\hookrightarrow$ i.e., the linear regression fit the conditional mean.

Bayesian quantile regression

In many cases it is useful to exploit X to understand the full distribution of y. A structured approach is to model the conditional quantiles of y,

$$Q_p\left(y|X\right) = X\beta_p, \tag{7}$$

where $p \in (0,1)$ denotes the quantile of y, and β_p corresponds to the regression parameters that correspond to the quantile p.

While the conditional quantile can be modelled using either non-linear or linear methods, the latter is certainly the most commonly used,

$$y = X\beta_p + \varepsilon, \tag{8}$$

Bayesian quantile regression

Koenker and Bassett Jr (1978) showed that the parameters β_p in the linear specification in Eq.(8) can be estimated as

$$\widehat{\beta}_p = \min_{\beta_p} \mathbb{E}\left(\sum_{i=1}^{n} \rho_p\left(\varepsilon_i\right)\right),$$

where $\rho_p\left(u\right) = \left(r - \mathbb{I}\left(u < r\right)\right)$ is a loss function, and $I\left(a\right)$ denotes an indicator function that takes value one if the event a is true, and zero otherwise.

In other words, β_p depends on the pth quantile of the random error term ε_i, which is defined as the value q for which $Pr\left(\varepsilon_i < q\right) = p$.[2]

[2] The distribution of the error terms is often left unspecified and is restricted to have the pth quantile equal to zero, i.e., $\int_{-\infty}^{0} f_p\left(\varepsilon_i\right) d\epsilon_i = p$.

Bayesian quantile regression

Following Yu and Moyeed (2001), we consider the linear model given by

$$y_i = \boldsymbol{x}_i^\top \beta_p + \varepsilon_i, \qquad i = 1, \ldots, n$$

and assume that ε_i has the asymmetric Laplace distribution with density

$$f_p\left(\varepsilon_i\right) = p(1-p) \exp\left\{-\rho_p\left(\varepsilon_i\right)\right\},$$

where $\rho_p(\cdot)$ is defined above. It is know that the mean and the variance of the asymmetric Laplace distribution are given by,[3]

$$E\left(\varepsilon_i\right) = \frac{1-2p}{p(1-p)} \qquad \text{and} \qquad Var\left(\epsilon_i\right) = \frac{1-2p+2p^2}{p^2(1-p)^2},$$

[3]Some other properties of the asymmetric Laplace can be found in Yu and Zhang (2005).

Bayesian quantile regression

Starting from the linear representation above, Kozumi and Kobayashi (2011) show that the conditional likelihood admits a mixture representation based on a scaled exponential normal distribution of the error term

$$\varepsilon = \theta z + \tau \sqrt{z} u \tag{9}$$

where

$$\theta = \frac{1-2p}{p(1-p)} \qquad \text{and} \qquad \tau^2 = \frac{2}{p(1-p)}$$

From this result, the linear model can be equivalently rewritten as

$$y = X\beta_p + \theta z + \tau \sqrt{z} u \tag{10}$$

where $z \sim \mathrm{Exp}(1)$ and $u \sim N(0,1)$ are mutually independent.

Bayesian quantile regression

As the conditional distribution of y given z is normal with mean $X\beta_p + \theta z$ and variance $\tau^2 z$, the joint density of y is given by

$$y|\beta_p, z \propto \left(\prod_{i=1}^{n} z_i^{-1/2}\right) \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} \frac{\left(y_i - \boldsymbol{x}_i^\top \beta_p - \theta_p z_i\right)^2}{2\tau^2 z}\right\}, \tag{11}$$

To proceed with the Bayesian estimation, Kozumi and Kobayashi (2011) assume the prior

$$\beta_p \sim N\left(\beta_0, B_0\right),$$

where β_0 and B_0 are the prior mean and covariance of β_p.

Yu and Moyeed (2001) proved that all posterior moments of β_p exist under such normal prior.

Bayesian quantile regression

A Gibbs sampling algorithm can be used to sample $\beta_p|y, z \sim N\left(\widehat{\beta}_p, \widehat{B}_p\right)$, where

$$\widehat{B}_p^{-1} = \sum_{i=1}^{n} \frac{\boldsymbol{x}_i \boldsymbol{x}_i^\top}{\tau^2 z_i}, \qquad \text{and} \qquad \widehat{\beta}_p = \widehat{B}_p \left\{ \sum_{i=1}^{n} \frac{\boldsymbol{x}_i \left(y_i - \theta z_i\right)}{\tau^2 z_i} + B_0^{-1} \beta_0 \right\},$$

From Eq.(11), one can show that the full conditional of z is proportional to

$$z_i^{-1/2} \exp\left\{ -\frac{1}{2} \left(\widehat{\kappa}_{1i}^2 z_i^{-1} + \widehat{\kappa}_{2i}^2 z_i\right) \right\}, \quad \text{where} \quad \widehat{\kappa}_{1i}^2 = \left(y_i - \boldsymbol{x}_i^\top \beta\right)^2 / \tau^2 \quad \text{and} \quad \widehat{\kappa}_{2i}^2 = 2 + \theta^2/\tau^2$$

This is the kernel of a generalised inverse Gaussian so that

$$z_i|y, \beta_p \sim \text{GIG}\left(\frac{1}{2}, \widehat{\kappa}_{1i}^2, \widehat{\kappa}_{2i}^2\right),$$

Extensions to large dimensions

Korobilis (2017) examines the role of model uncertainty in quantile forecasts by coupling the linear specification outlined above with a spike-and-slab prior on the regression coefficients β_p.

Specifically, Korobilis (2017) proposed a hierarchical prior of the form

$$\beta_{ip}|\gamma_{ip}, \delta_{ip} \sim (1-\gamma_{ip})\, N\left(0, c \times \delta_{ip}^2\right) + \gamma_{ip} N\left(0, \sigma_{ip}^2\right),$$

with

$$\delta_{ip}^{-2} \sim G\left(a_1, a_2\right), \qquad \gamma_{ip}|\pi_0 \sim Ber\left(\pi_0\right), \qquad \pi_0 \sim Beta\left(b_1, b_2\right),$$

where $c \to 0$ is a fixed parameter. When $\gamma_{ip} = 1$, β_{ip} has a normal prior with variance δ_{ip}^2. When $\gamma_{ip} = 0$, β_{ip} has a normal prior with variance $c \times \delta_{ip}^2$, which will be very close to zero for c small enough.

Extensions to large dimensions

Korobilis (2017) shows that draws from the posterior distribution $\beta_p|\gamma_p, \tau^2, y, z$ are obtained by sampling sequentially from,

$$\beta_p|\gamma_p, \tau^2, y, z \sim N\left(\widehat{\beta}_p, \widehat{B}_p\right),$$

where

$$\widehat{B}_p^{-1} = \left(\sum_{i=1}^{n} \frac{\tilde{\boldsymbol{x}}_i^\top \tilde{\boldsymbol{x}}_i}{\tau^2 z_i} + \Delta_p^{-1}\right), \qquad \widehat{\beta}_p = \widehat{B}_p \left[\sum_{i=1}^{n} \frac{\tilde{\boldsymbol{x}}_i \left(y_i - \theta z_i\right)}{\tau^2 z_i}\right],$$

and Δ_p is a diagonal matrix with elements δ_{ip}^2 if $\gamma_{ip} = 1$ and $c \times \delta_{ip}^2$ if $\gamma_{ip} = 0$.

Extensions to large dimensions

Similarly, the posterior from the remaining parameters of the mixture prior can be sampled from

$$\delta_{ip}^{-2}|\beta_{ip}, y \sim G\left(\widehat{a}_1, \widehat{a}_2\right), \quad \text{where} \quad \widehat{a}_1 = a_1 + \frac{1}{2}, \quad \widehat{a}_2 = a_2 + \frac{\beta_{ip}^2}{2},$$

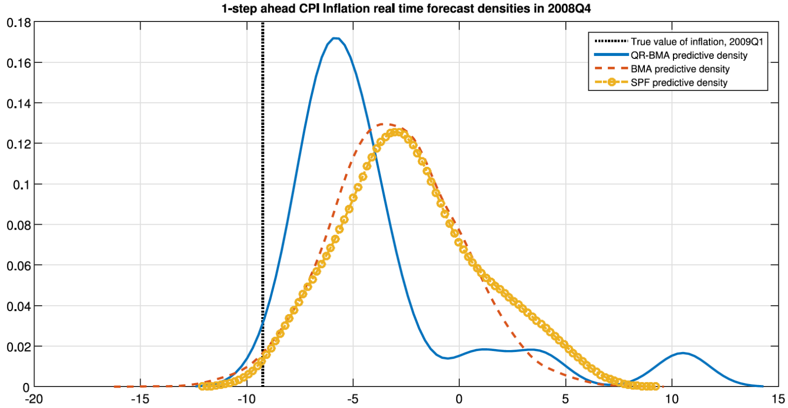
and $\gamma_{ip}|\gamma_{-/i,p}, \beta_{ip}, z, y \sim \text{Ber}\left(\overline{\pi}\right)$, where

$$\overline{\pi} = \frac{\pi_0 p\left(\gamma_{ip} = 1|\gamma_{-/ip}, \tilde{y}\right)}{\pi_0 p\left(\gamma_{ip} = 1|\gamma_{-/ip}, \tilde{y}\right) + (1-\pi_0)\pi_0 p\left(\gamma_{ip} = 0|\gamma_{-/ip}, \tilde{y}\right)},$$

with $\tilde{y} = y - \theta z$, $\gamma_{-/ip}$ denotes the vector γ_p without the ith element removed, and $p\left(\gamma_{ip} = j|\gamma_{-/ip}, \tilde{y}\right)$ the likelihood of $\tilde{y}_i = y_i - \theta z_i = \boldsymbol{x}_i^\top \beta_p + \tau\sqrt{z_i}u_i$ evaluated assuming $\gamma_{ip} = j$ for $j = 1, 0$.

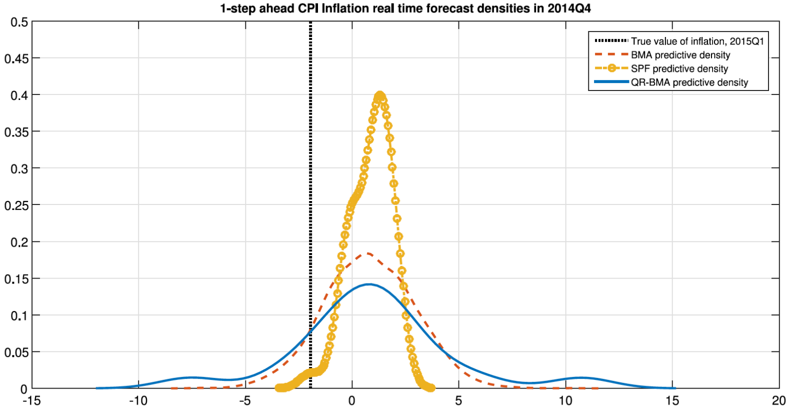
Finally, we sample π_0 from $\pi_0|\gamma_p, \beta_p, z, y \sim \text{Beta}\left(\widehat{b}_1, \widehat{b}_2\right)$ where $\widehat{b}_1 = p_\gamma + b_1$ and $\widehat{b}_2 = p - p_\gamma + b_2$ and $p_\gamma = \sum_i \gamma_{ip}$.

Extensions to large dimensions



Predictive densities of inflation. Source: Korobilis (2017).

Extensions to large dimensions



Predictive densities of inflation. Source: Korobilis (2017).

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679.

Bianchi, D. and Chiarella, C. (2019). An anatomy of industry merger waves. *Journal of Financial Econometrics*, 17(2):153–179.

Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316.

Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika*, 93(4):827–841.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies*, 65(3):361–393.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Korobilis, D. (2017). Quantile regression forecasts of inflation under model uncertainty. *International Journal of Forecasting*, 33(1):11–20.

Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation*, 81(11):1565–1578.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447.

Yu, K. and Zhang, J. (2005). A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics‚ÄîTheory and Methods*, 34(9-10):1867–1879.