CROATIAN NATIONAL BANK

Working Papers W-30

Estimating Credit Migration Matrices with Aggregate Data – Bayesian Approach

Davor Kunovac

Zagreb, November 2011

WORKING PAPERS W-30

PUBLISHER
Croatian National Bank
Publishing Department
Trg hrvatskih velikana 3, 10002 Zagreb
Phone: +385 1 45 64 555
Contact phone: +385 1 45 65 006
Fax: +385 1 45 64 687

WEBSITE
www.hnb.hr

EDITOR-IN-CHIEF
Evan Kraft

EDITORIAL BOARD
Ljubinko Jankov
Gordi Sušić
Maroje Lang
Boris Vujčić

EDITOR
Romana Sinković

ENGLISH LANGUAGE EDITOR
Prevoditeljski centar d.o.o.

DESIGNER
Vjekoslav Gjergja

TECHNICAL EDITOR
Slavko Križnjak

PRINTED BY
Printera Grupa d.o.o.

Printed in 300 copies

ISSN 1331-8586 (print)
ISSN 1334-0131 (online)

CROATIAN NATIONAL BANK

WORKING PAPERS W-30

Estimating Credit Migration Matrices with Aggregate Data – Bayesian Approach

Davor Kunovac

Zagreb, November 2011

Abstract

This paper studies the Gibbs sampler developed in Rodriguez-Yam et al. (2004) that can be used to estimate the parameters of inequality constrained regression. The aim of the paper is twofold. First, to present an efficient estimation methodology that has not yet been utilised in credit risk literature, and second, by using this method to estimate a migration matrix governing the dynamics of loans to corporate sector in Croatia when only aggregate (*supervisory*) data on bank loans in every rating category is available. The results of the analysis suggest that the Bayesian estimator used in this paper has some important comparative advantages over the estimators previously used in the literature.

JEL:
C13, C50

Keywords:
credit migration matrices, Bayesian inference, inequality constrained regression, truncated normal vector

I would like to thank Gordon Kemp, Bojan Basrak and Evan Kraft whose comments on earlier drafts have improved the paper. I would also like to thank Ivan Huljak for his help with data.

Contents

1 Introduction and motivation

1.1 Preliminary discussion

The aim of the paper is twofold. Firstly, to present and test an efficient methodology to estimate the parameters of the inequality constrained regression that has not yet been used in credit risk literature. Secondly, by using this method to estimate a migration matrix governing dynamics of loans to corporate sector in Croatia when only aggregate (*supervisory*) data on bank loans in every quality state (*rating*) is available. The empirical analysis conducted in this paper therefore deals with the modelling of migration of bank loans between different quality classes. In particular, the central question to be answered is: what is the probability for a *good* loan to become a *bad* (or non performing) loan and vice versa?

Migration matrices are standardly used to describe the dynamics of a credit portfolio over time and are often seen as a useful tool when forecasting credit quality. In a case in which data on transition between ratings for individual loans is available, the transition probabilities are easily calculated via maximum likelihood from relative frequencies of *arrivals and departures* for each rating. However, only aggregate time series of credit volume in each quality class, usually compiled by the regulator, is typically available. In that case, migration probabilities may be seen as transition probabilities of a first order, discrete time and discrete state space Markov chain which casts the estimation problem into the inequality constrained regression framework (Kelton 1981, Lee et al. 1965, Rodriguez-Yam et al. 2004). Literature that exploits these minimal data requirements when estimating migration matrices is, however, not abundant. For example, Jones (2005) estimates transition matrices for the US bank loan portfolio using the *Inequality Constrained Least Squares Estimator*. The absence of any relevant asymptotic theory for this estimator and its boundary solutions complicates the inference. Transition probabilities are, for example, often estimated as zeros, which implies extreme conclusions that may be misleading and hard to interpret. In order to circumvent the problems of the least squares estimates, Christodoulakis (2006) studies the same data set, but now using a different estimation strategy. The transition probabilities are estimated under Monte Carlo integration framework, as proposed by Kloek and van Dijk (1978) and Geweke (1986). Empirical results on the US portfolio of non-performing loan proportions reported in Christodoulakis (2006) are in some cases close to the estimates of Jones (2005), but also exhibit some statistically significant differences regarding the estimated transition probabilities. However, sampling based on the Monte Carlo integration may be slow and inefficient. Christodoulakis (2006) does not report on performances of his sampler, but only that the inference is based on a large number of replications (100000).

This paper estimates transition probabilities for a typical loan to the corporate sector in Croatia. It builds on the work by Jones (2005) and Christodoulakis (2006) by employing an alternative, Bayesian, strategy when estimating transition probabilities between quality classes.

1.2 Inequality constrained regression

Under the assumption that credit quality of a typical loan obeys a first order Markov Chain, the related transition/migration matrix is estimated under the inequality constrained regression framework. For that reason the estimation of this model is now discussed in more details.

In practice, econometricians often face the problem of estimating a linear regression model where some of the parameters are known to belong to *a priori* given intervals. The model is defined by a set of inequality restrictions imposed on the coefficients of the classical regression model and is referred to as the *inequality constrained regression model*. The prior knowledge about the parameters mainly relies on the requirements set by economic or statistical theory. Alternatively, inequality constraints just represent *ad hoc* restrictions, considered to be *self-evident* from the nature of the problem of interest.

The most common set of constraints met in practice are *the sign restrictions* imposed on the regression parameters. The theory often presumes the signs of marginal effects of individual regressors in the model. The signs of estimated coefficients, however, need not to be consistent with those anticipated by the theory. For example, the sign inconsistency often arises due to highly correlated regressors that may mix individual effects in the model. As reported in Geweke (1986), procedures used to deal with this practical problem differ from one investigator to another. Typically, regressors that produce the *wrong* sign are simply dropped out of the specification and only results based on specifications with all the expected signs are reported. However, such a practice may result in severely biased estimates and can lead to inefficient parameter estimates (Lovell and Prescott 1970). On the other hand, simply dropping a variable out of the linear model is not always an available option, particularly in a case in which the underlying theoretical model originates from the probability theory. To illustrate, it has been long recognised that the estimation of transition probabilities governing a discrete Markov Chain easily reduces to the problem of the estimation of an inequality constrained regression (Lee et al. 1965, MacRae 1977). Clearly, the elements in each row of the *transition matrix* must be non-negative, less or equal than one and need to sum up to unity. Exactly the same set of restrictions must therefore arise in the pertaining regression.

Instead of searching for a set of explanatory variables that finally delivers parameters with expected signs, an alternative approach considers the estimation of a restricted model where all the parameters are *forced* to produce expected signs, or more generally, to satisfy given inequality constraints. Following this idea, Judge and Takayama (1966) solve a quadratic programme via the *Dantzig-Cottle* algorithm and minimise the residual sum of squares of a linear model under inequality restrictions on the parameters. In this way the *Inequality Constrained Least Squares Estimator (ICLS)* is obtained. There are several major problems related to this approach. The asymptotical properties of the ICLS are clear and well documented for several simple cases only and, accordingly, are not particularly useful in practice. The standard test for significance of coefficients based on Student distribution can be misleading (Lovell and Prescott 1970) and moments of the estimator are difficult to derive (Judge and Takayama 1966). From a more practical point of view, not all the standard econometric software packages include built-in procedures that easily produce inequality constrained estimators. Finally, a solution obtained via quadratic programming typically lies on the boundary of the feasibility region and, hence, is often considered as being non-informative.

In order to circumvent the shortcomings of the ICLS estimator, the problem has also been approached from a Bayesian perspective. Early theoretical work includes O'Hagan (1973), Davis (1978) and Chamberlain and Leamer (1976). However, only after the popularisation of computationally demanding Monte Carlo methods during the 1980s and 1990s did Bayesian inequality constrained estimators become relevant for practitioners. Building on the previous work by Kloek and van Dijk (1978), Geweke (1986) draws a sample from posterior distributions of inequality constrained regression parameters using the Monte Carlo integration. However, the methodology proved to be inefficient and slow to the point of impracticality in a case in which the parameter space has a *small* constrained region and/or the number of parameters to be estimated is *large* (Geweke 1996). The *Gibbs sampler* aimed to simulate from posterior for inequality constrained regression parameters was first presented in Geweke (1996). The prior distributions used were diffuse priors that are standardly used in linear Bayesian models, but now constrained in order to reflect the prior knowledge about

parameters. Posterior distribution for coefficients in this case takes on the form of truncated multivariate normal distribution and the problem of sampling from the posterior collapses to sampling from truncated normal distribution. Sampling from multivariate truncated distributions, however, is not a trivial task, except for low dimensions. For that purpose, Geweke uses a methodology described in Hajivassiliou and McFadden (1990), Hajivassiliou, McFadden, and Ruud (1990) and Geweke (1991). Although Geweke's sampler is suitable for solving a broad class of problems, many drawbacks remain. Most importantly, the number of restrictions that can be imposed on regression coefficients must not exceed the overall number of parameters to be estimated. Moreover, restrictions need to be linearly independent. For example, these shortcomings do not allow one to apply this sampler when estimating transition probabilities of a Markov Chain under a constrained regression framework. In addition, Geweke's sampler produces highly autocorrelated draws and mixes poorly, i.e. explores the posterior distribution inefficiently. Of course, posterior draws from a Gibbs sampler are realisations of a Markov Chain and, as such, are correlated by definition. Still, the problem here is serious and causes slow convergence toward the posterior distribution.

Finally, the Gibbs sampler constructed in Rodriguez-Yam, Davis and Scharf (2004) overcomes the problems that characterise earlier specifications. The authors illustrate that their sampler can incorporate more restrictions than Geweke's (1996), it mixes better, produces less correlated draws and converges toward posterior distribution faster than competitive implementations. For this reason, this paper studies the sampler of Rodriguez-Yam, Davis and Scharf (2004) in more detail.

The estimation strategy adopted here relies on Rodriguez-Yam et al. (2004). Probabilities of migration between quality classes for Croatian banks' loan portfolios are estimated efficiently via the Gibbs sampler, where the convergence of Bayesian estimators is achieved after less than 10000 iterations. In addition, estimated probabilities lie well within the truncation area producing estimates that have clear interpretation.

The rest of the paper is structured as follows. The following section illustrates how one can estimate a credit migration matrix with aggregate data. After that, the third section lays out technical details on the sampler I rely on throughout the paper. In order to be able to judge the reliability of the estimator used in this paper the fourth section additionally compares its properties with those of the OLS and ICLS in a simple Monte Carlo experiment. The fifth section reports the estimation results and the final section concludes.

2 Estimating credit migration matrices with aggregate data

This section demonstrates how and under what conditions the estimation of transition probabilities governing a discrete Markov chain reduces to the estimation of an inequality constrained regression.

As a part of the risk management process, financial institutions often construct simple econometric models in order to summarise past behaviour of various financial indicators. Hopefully, such information will be useful for forecasting and, consequently, may provide individual institutions, as well as their regulators, with an important input while they are managing credit risk.

The basic building block of many credit risk models is *the migration matrix* of a loan portfolio. Its purpose is to summarise the probabilities of the migration of loans in a portfolio between different ratings. For example, it is naturally of interest to risk managers to estimate the probability of a good (performing) loan to become a bad (non-performing) loan or vice versa. In a case in which the mechanism underlying migration between ratings remains similar in the following periods, migration matrices are a powerful tool for the prediction of bank losses.

2.1 Transition matrix of a Markov chain under regression framework

Often, it is assumed that the transition of loans between different ratings obeys a discrete Markov chain process. When only aggregate data on the proportions of loans in every risk category is available, the direct calculation of transition probabilities is not possible and the problem of estimation the migration matrix is cast into the inequality constrained regression framework (Lee et al. 1965, Rodriguez-Yam et al. 2004) as follows.

Let $S = \{S_1, \ldots, S_k\}$ denote the discrete state space (of loan ratings) of a time homogenous[1] Markov chain $X = \{X_0, X_1, \ldots\}$. The problem of interest here is that of the estimation of the transition probabilities between the states summarised by $k \times k$ matrix P such that:

$$P = [p_{ij}]_{i,j=1}^{k} = [P(X_t = S_j \,|\, X_{t-1} = S_i)]_{i,j=1}^{k}.$$

The estimation strategy depends on the nature of the available data. In cases when the outcomes of the chain X are observed directly, transition probabilities are easily obtained via maximum likelihood as (see Lee et al. 1968):

$$P_{ML} = [\hat{p}_{ij}]_{i,j=1}^{k} = \left[\frac{n_{ij}}{\sum_{j=1}^{k} n_{ij}}\right]_{i,j=1}^{k},$$

where n_{ij} denotes the number of observed migration from state i to state j. We refer to this approach as to the *micro-data* estimation. More often, the outcomes of the chain are not observable directly, but only the aggregate number of units in each state of the chain over time. These units are assumed to be mutually independent and follow the same Markov process with transition matrix P. Individual transitions are not observed in this *macro-data* case and a different estimation strategy has to be adopted.

For that purpose, let us first evaluate the joint probability for two sequential realisations of a chain

$$P(X_{t-1} = S_i, X_t = S_j) = P(X_{t-1} = S_i)P(X_t = S_j \,|\, X_{t-1} = S_i) = P(X_{t-1} = S_i)p_{ij}. \tag{1}$$

The unconditional probability for the chain at time t to be in state j is now:

$$P(X_t = S_j) = \sum_{i=1}^{k} P(X_{t-1} = S_i, X_t = S_j) = \sum_{i=1}^{k} P(X_{t-1} = S_i)p_{ij}. \tag{2}$$

In practice, probabilities $P(X_t = S_j)$ are estimated as the proportions of units being in state j at time t, denoted y_{jt}. Clearly, these proportions are easily obtained from the aggregate data set (macro data), even though the outcomes of the underlying chain are not observed. As a consequence, the sample counterpart of (2) casts into the inequality constrained regression framework with the following specification to be estimated:

$$y_{jt} = \sum_{i=1}^{k} y_{it-1}p_{ij} + \varepsilon_{jt}, \tag{3}$$

imposing

$$\sum_{j=1}^{k} p_{ij} = 1 \text{ for all } i \tag{4}$$

$$p_{ij} \geq 0 \text{ for all } i, j, \tag{5}$$

where $i = 1,\ldots,k$ are states of a Markov chain (i.e. possible ratings of a loan), and y_{jt} is the proportion of loans of quality j at time t. Moreover, a single absorbing (terminal) state induces additional k restrictions:

1 Time homogeneity of a chain reflects the fact that transition probabilities do not vary over time.

$$p_{kk} = 1 \text{ and } p_{k1} = p_{k2} = \ldots = p_{kk-1} = 0. \tag{6}$$

Equality constraints are included into the specification in a standard way. Plugging (4) into (3) and assuming a single absorbing state (6) yields *the reduced model* which can be estimated via the inequality constrained estimator. However, equality constraints here impose cross-equation restrictions onto regression parameters in (3) and, accordingly, regressions for all j need to be estimated simultaneously. For our purpose, it is sufficient to illustrate the procedure for a four-state chain (i.e. $k = 4$ with a single absorbing state.

Let us consider a set of equations (3) under (6) but now stacked as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \Delta y_4 \end{pmatrix} = \begin{pmatrix} X & 0 & 0 & 0 \\ 0 & X & 0 & 0 \\ 0 & 0 & X & 0 \\ 0 & 0 & 0 & X \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} + \varepsilon,$$

where $y_j = (y_{j2}, \ldots, y_{jT})^\tau$, X is common matrix of regressors from (3), p_j denote columns of transition matrix, Δ is standard differencing operator and $\varepsilon \sim N(0, \sigma^2 I)$ is error term.[2]

In order to proceed with the estimation, let us partition this stacked matrix of regressors in four parts:

$$X_1 = \begin{pmatrix} X \\ 0 \\ 0 \\ 0 \end{pmatrix}, X_2 = \begin{pmatrix} 0 \\ X \\ 0 \\ 0 \end{pmatrix}, X_3 = \begin{pmatrix} 0 \\ 0 \\ X \\ 0 \end{pmatrix} \text{ and } X_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ X \end{pmatrix}.$$

Now, for all i we can rewrite (4) as:

$$p_{i4} = 1 - p_{i1} - p_{i2} - p_{i3}, \tag{7}$$

or $p_4 = \mathbf{1} - p_1 - p_2 - p_3$ in matrix notation, which yields:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \Delta y_4 \end{pmatrix} = (X_1 X_2 X_3 X_4) \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ \mathbf{1} - p_1 - p_2 - p_3 \end{pmatrix} + \varepsilon = X_1 p_1 + X_2 p_2 + X_3 p_3 + X_4 (\mathbf{1} - p_1 - p_2 - p_3) + \varepsilon.$$

Finally, we have

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \Delta y_4 \end{pmatrix} - X_4 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = (X_1 - X_4) p_1 + (X_2 - X_4) p_2 + (X_3 - X_4) p_3 + \varepsilon \tag{8}$$

which is an equation that can be estimated under the inequality constrained framework by imposing:

$$p_{ij} \geq 0, \text{ for } i = 1,2,3 \text{ and } j = 1,2,3 \tag{9}$$

$$p_{i1} + p_{i2} + p_{i3} \leq 1, \text{ for } i = 1,2,3. \tag{10}$$

2 The fact that we are dealing with proportions data implies a non-zero contemporaneous correlation between the error terms across equations. Indeed, adding up both sides in $y_{jt} = \sum_{i=1}^{K} y_{it-1} p_{ij} + \varepsilon_{jt}$ over $j = 1,2,3,4$ taking into account the constraints on the transition probabilities as well as $\sum_{j=1}^{4} y_{jt} = 1$ reveals the link between errors as $\sum_{j=1}^{4} \varepsilon_{jt} = 0$ for all t. As a consequence, the least squares yield *unbiased and consistent* (Lee, Judge and Zellner 1970), but *not efficient* estimates of transition probabilities. Because of that, MacRae (1977) proposes an iterative GLS estimator. Finally, Kelton (1981) shows the equivalence of the GLS and least squares estimators under mild conditions. The main purpose of this example is to compare the estimates based on the sampler from Rodriguez-Yam, Davis and Scharf (2004) with more conventional solutions obtained via OLS and ICLS and, therefore, despite the non-optimality of the choice, the covariance matrix is assumed to be diagonal, i.e. $\varepsilon \sim N(0, \sigma^2 I)$.

As noted in Rodriguez-Yam et al. (2004), the number of restrictions imposed here is greater than the number of parameters to be estimated and, therefore, the alternative Gibbs sampler (Geweke 1996) is not an appropriate estimation method.

3 Bayesian estimation of inequality constrained regression

This section presents the Bayesian estimator for the inequality constrained linear regression proposed by Rodriguez-Yam, Davis and Scharf (2004). This sampler, based on posterior draws simulated via Gibbs sampler, overcomes the main problems of the earlier specifications. The authors illustrate that their sampler can easily incorporate a large number of, possibly, collinear restrictions, it mixes better, produces less correlated draws and converges toward posterior distribution faster than competitive Bayesian implementations. In contrast to that, the non-Bayesian or classical approach appears to be problematic due to the absence of any relevant asymptotic theory for inequality constrained estimators. Moreover, the boundary solutions produced in this way may be seriously misleading.

Bayesian normal linear regression model standardly assumes the normal prior for regression coefficient and, consequently, due to the form of the likelihood function, the posterior appears to be normally distributed as well. The same applies for constrained regression, but now with an important difference. The prior knowledge about the coefficients is built into the model by using the truncated normal distribution as the prior and, consequently, the posterior is again from the truncated normal distribution. In contrast to the unrestricted case, theoretical moments of restricted Gaussian distribution are generally not derived easily and, thus, MCMC methods are used to sample from the posterior. Still, as mentioned before, sampling from truncated multivariate distribution is not a trivial task either and the estimator for the constrained regression model presented in Rodriguez-Yam, Davis and Scharf (2004) is, in fact, an application of an efficient Gibbs sampler from the truncated normal distribution.

The following section firstly presents the Gibbs sampler for truncated normal distribution and, after that, illustrates how the sampler fits naturally into inequality constrained regression framework.

3.1 Sampling from truncated normal distribution

A truncated normal vector is formally defined in the following definition (Rodriguez-Yam, Davis and Scharf 2004).

Definition 1 (Truncated normal vector) *If R is a subset of $\mathbb{R}^k$ having positive Lebesgue measure, we call the random k-vector Y truncated normal and write $Y \sim N_R(\mu, \Sigma)$ if its probability density function is proportional to $f(x;\mu,\Sigma)I_R(x)$, where $f(x;\mu,\Sigma)$ is the k-variate normal density with mean μ and covariance matrix Σ, and $I_R(x)$ is the indicator function for R.*

The Gibbs algorithm for truncated normal distribution relies on some good properties that characterise linear transformations and conditional distributions of the unconstrained normal vector, also inherited by constrained vector. These standard results ensure that a linear transformation of a constrained normal vector is normally distributed as well, and, also, the normality of univariate conditionals is guaranteed. Recalling that the Gibbs sampler draws rely on the sampling from conditional distributions, the importance of these results becomes apparent. More formally, these results can be stated in form of the two propositions (see appendix).

The final objective here is to simulate a random sample from a vector

$$X \sim N_T(\mu, \sigma^2\Sigma), \text{ where } T = \left\{x \in \mathbb{R}^k : Bx \leq b\right\}.$$

In contrast to similar applications, matrix B is allowed to have collinear rows here, i.e. inequality constraints do not need to be linearly independent. Let A be a square matrix of full rank, such that

$$A\Sigma A^\tau = I_k \tag{11}$$

and let $Z = AX$ be a linear transformation of X. Proposition 1 (see appendix) and (11) imply the distribution of Z:

$$Z \sim N_S(A\mu, \sigma^2 A\Sigma A^\tau) = (2) = N_S(A\mu, \sigma^2 I_k), \text{ where } S = \{Ax : x \in \mathbb{R}^k \;\&\; Bx \leq b\}. \tag{12}$$

The area S may be rewritten as

$$S = \{z \in \mathbb{R}^k : Dz \leq b\}, \text{ where } D = BA^{-1}. \tag{13}$$

The purpose of the transformation of vector X in (12) is the simplification of covariance matrix. As a consequence, all the univariate conditionals needed to implement one round of the Gibbs sampler are defined fully provided μ and σ. What remains is to deal with truncation sets for univariate conditional distribution of vector Z. For that purpose, let us define:

$$\alpha = A\mu, Z_{-j} = (Z_j, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_k) \text{ and } z_{-j} = (z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_k),$$

Z_{-j} denoting vector Z with its jth component excluded and z_{-j} is a real vector in $\mathbb{R}^k$. Proposition 2 implies that:

$$Z_j \big| (Z_{-j} = z_{-j}) \sim N_{S_j}(\alpha_j, \sigma^2), \tag{14}$$

where

$$S_j = \{z_j \in \mathbb{R} : z \in S\} = \{z_j \in \mathbb{R} : Dz \leq b\}.$$

Let D = $(d_1,\ldots,d_k)$ be the column representation of D and D_{-j} a matrix when jth column is deleted from D. Now, the set S_j may be conveniently rewritten as

$$S_j = \{z_j \in \mathbb{R} : d_j z_j \leq b - D_{-j} z_{-j}\}. \tag{15}$$

All the objects D_{-j}, z_{-j}, d_j and b are known and, therefore, due to convexity of truncation area, the boundaries of S_j may be found from the set of linear inequalities $d_j z_j \leq b - D_{-j} z_{-j}$. After deriving the boundaries of S_j in (15), one can easily sample from univariate conditionals $Z_j \big| (Z_{-j} = z_{-j})$ given in (14). Although the algorithm provides a draw from transformed vector $Z = AX$, premultiplying this with the matrix A^{-1} yields a draw from X. The entire procedure is easily implemented by following the steps of the algorithm:

Algorithm (Gibbs sampler for truncated multivariate normal vector)

1. Initialise the chain with $z_0 \in S$. Suppose that in t-th step we have $z^{(t)} = (z_j^{(t)}, \ldots, z_k^{(t)})$
2. Simulate $z_j^{(t+1)}$ from $f(z_j \big| z_1^{(t+1)}, \ldots, z_{j-1}^{(t+1)}, z_{j+1}^{(t)}, \ldots, z_k^{(t)})$, for $j = 1,\ldots,k$
3. Undo the linear transformation: $X^{(t+1)} = A^{-1} z^{(t+1)}$.

3.2 Gibbs sampler for inequality constrained regression

This section introduces the Gibbs sampler aimed at simulating from the posterior distribution of parameters of the inequality constrained regression. Simulation from the posterior of the regression coefficients reduces to simulating from truncated normal distribution. For that purpose, the Gibbs scheme introduced in the previous section is utilised.

The basic model we consider is a k-variate regression, where regression coefficients need to satisfy a set of linear inequality constraints. In matrix form we write this *inequality constrained regression* as follows

$$Y = X\beta + \varepsilon, \text{ such that } B\beta \leq b,$$

where $\varepsilon \sim N(0, \sigma^2 I_n)$ is a standard noise vector B denotes a known $l \times k$ matrix of constraints, where l is allowed to be greater than k and b is a known k-vector. Inequality constraints define a set:

$$T = \{\beta \in \mathbb{R}^k : B\beta \leq b\}.$$

The set of parameters to be estimated includes regression coefficients β and noise variance σ. Jointly they form a $k+1$-vector:

$$\theta = (\beta, \sigma).$$

The estimation framework is a Bayesian and, therefore, the model is defined fully by defining the prior distribution for θ. A standard choice for Bayesian linear models is to set a normal prior for β and an *inverse-gamma*[3] prior for σ.

$$\beta \sim N_T(\mu_0, \sigma_0^2 (X^\tau X)^{-1}) \tag{16}$$

$$\sigma^2 \sim IG(\nu, \lambda), \tag{17}$$

where σ_0, ν and λ denote known scalars and μ_0 a known vector. Suppose in addition that β and σ are independent. The prior knowledge about the coefficients is built into the model by using the truncated normal distribution as the prior and, consequently, the posterior is again from the truncated normal distribution. In contrast to the unrestricted case, the moments from restricted Gaussian distribution are generally not derived easily and, thus, MCMC methods are used to sample from the posterior.

In order to derive the posterior for $\theta = (\beta, \sigma)$ we use the Bayes theorem:

$$f(\beta, \sigma^2 | y) = \frac{f(y | \beta, \sigma^2) f(\beta, \sigma^2)}{f(y)} \propto f(y | \beta, \sigma^2) f(\beta, \sigma^2) = L(\beta, \sigma^2 | y) f(\beta) f(\sigma^2). \tag{18}$$

where $L(\beta, \sigma^2 | y)$ is the standard Gaussian likelihood function:

$$L(\beta, \sigma^2 | y) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-(y - X\beta)^\tau (y - X\beta)/2\sigma^2\}. \tag{19}$$

Now, using expressions (16), (17), (18) and (19) the following posteriors can be derived:

$$\beta | (\sigma^2, y) \sim N_T(\mu_1, \Sigma_1) \tag{20}$$

$$\sigma^{-2} | (\beta, y) \sim (SS(\beta) + 2\lambda^{-1})^{-1} x^2_{n+2\nu}, \tag{21}$$

3 A random variable X is said to be inverse-gamma distributed with parameters a and b, $X \sim IG(a, b)$, if $\frac{1}{X}$ is gamma distributed with parameters a and b.

where:

$$\mu_1 = \gamma\hat{\beta} + (1-\gamma)\mu_0 \tag{22}$$

$$\Sigma_1 = \sigma^2\gamma(X^\tau X)^{-1} \tag{23}$$

$$SS(\beta) = (y - X\beta)^\tau (y - X\beta) \tag{24}$$

$$\gamma = \sigma_0^2/(\sigma_0^2 + \sigma^2) \tag{25}$$

$$\hat{\beta} = (X^\tau X)^{-1} X^\tau y. \tag{26}$$

While normal posterior for β is a standard result in literature, the $x^2_{n+2\nu}$ posterior for variance may be less obvious to derive (see appendix).

Vector $\mu_1 = \gamma\hat{\beta} + (1-\gamma)\mu_0$ is a convex combination of the prior parameter μ_0 and the OLS estimator $\hat{\beta}$. How much weight is attached to individual vector μ_0 or $\hat{\beta}$ depends on the uncertainty on the prior reflected in $\gamma = \sigma_0^2/(\sigma_0^2 + \sigma^2)$. For example, if there is a high level of certainty about the prior distribution, variance of the prior will be smaller, i.e. constant σ_0^2 will be small. As a consequence, γ will be closer to zero and the convex combination μ_1 thus closer to μ_0.

In a case in which the OLS estimator satisfies all the constraints, it is equal to the ICLS. Moreover, if the prior parameter μ_0 is set to be the ICLS, i.e. to the OLS now, it follows that $\mu_1 = \gamma\hat{\beta} + (1-\gamma)\mu_0 = \hat{\beta}$, and posterior takes on the form $\beta|(\sigma^2, y) \sim N_T(\hat{\beta}, \Sigma_1)$. As a consequence, the Bayesian point estimator will be *close* to the OLS.

In order to sample a draw from the posterior, in every round of the algorithm a k-vector from truncated multivariate distribution and a single value from $x^2_{n+2\nu}$ distribution are to be simulated. For that purpose, the strategy introduced in the previous section is adopted. Let A be a regular matrix such that $A(X^\tau X)^{-1}A^\tau = I_k$. In addition, let us transform the vector β using the convenient linear transformation

$$\eta = A\beta.$$

Now using proposition 2 and the previous discussion it follows that

$$\eta|(\sigma^2, y) \sim N_S(A\mu_1, \sigma^2\gamma A(X^\tau X)^{-1}A^\tau) = N_S(A\mu_1, \sigma^2\gamma I_k)$$

where

$$S = \{\eta \in \mathbb{R}^k : D\eta \leq b\}, \text{ where } D = BA^{-1}. \tag{27}$$

The Gibbs sampler is now easily implemented using the following algorithm.

Algorithm (Gibbs sampler for inequality constrained regression)

1. Initialise the chain with $\theta^{(0)} = (\eta_1^{(0)}, \ldots, \eta_k^{(0)}, \sigma^{2(0)})$. Suppose that $\theta^{(t)}$ is the last simulated value.
2. Value $\theta^{(t+1)} = (\eta_1^{(t+1)}, \ldots, \eta_k^{(t+1)}, \sigma^{2(t+1)})$
   - simulate $\eta_1^{(t+1)}$ from $f(\eta_1|\eta_2^{(t)}, \eta_3^{(t)}, \ldots, \eta_k^{(t)}, \sigma^{2(t)}, y)$,
   - simulate $\eta_2^{(t+1)}$ from $f(\eta_2|\eta_1^{(t+1)}, \eta_3^{(t)}, \ldots, \eta_k^{(t)}, \sigma^{2(t)}, y)$
   - $\vdots$
   - simulate $\sigma^{2(t+1)}$ from $f(\sigma^{2(t)})|\eta_1^{(t+1)}, \eta_2^{(t+1)}, \ldots, \eta_k^{(t+1)}, y)$.
3. Undo linear transformation: $\beta^{(t+2)} = A^{-1}\eta^{(t+1)}$
4. Repeat procedure for $\theta^{(t+2)}$, $\theta^{(t+3)}$,...

To summarise, an efficient method for sampling from posterior distribution of the parameters of inequality constrained regression is described in detail. The simulation is based on the Gibbs sampler which generates a Markov chain that converges in distribution towards posterior distribution and, accordingly, the simulated sequence approximates a random draw from posterior.

4 Micro vs macro-data estimation – simulated example

Often, only aggregate or *macro* data on proportions of visits to each state is available. In that case, the fact that estimated probabilities represent the migration probabilities of individual, independent *micro-units*, included in the macro sample, raises a question of the reliability of the estimated parameters. In practice, when micro-data is available, transition probabilities should always be estimated via maximum likelihood. On the other hand, when only aggregate data is available, it is difficult to reveal to what extent the estimates obtained via regression methods resemble structural transition probabilities, underlying the dynamics of micro-units. In order to address this issue a simple Monte Carlo experiment is conducted.

Using an arbitrary transition matrix:

$$P = \begin{pmatrix} 0.90 & 0.05 & 0.05 \\ 0.2 & 0.4 & 0.4 \\ 0.03 & 0.02 & 0.95 \end{pmatrix},$$

five independent chains of length n are simulated.[4] After that, from simulated micro-data, the time series of proportions of units in every state are derived in order to compare the estimates of transition probabilities based on macro and micro data. Table 1 compares the estimates obtained with different approaches. Finally, in order to illustrate the sensitivity of the estimates on the length of the series used, the exercise is conducted both for $n = 100$ and $n = 10000$.

Table 1 Transition probabilities

		N = 10000			N = 100		
		$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
MLE		0.90	0.05	0.05	0.90	0.06	0.04
OLS		0.90	0.05	0.05	0.87	0.05	0.08
ICLS	$i = 1$	0.90	0.05	0.05	0.87	0.05	0.08
Bayes		0.90	0.05	0.05	0.87	0.05	0.08
MLE		0.21	0.39	0.40	0.05	0.38	0.57
OLS		0.21	0.39	0.40	–0.10	0.39	0.71
ICLS	$i = 2$	0.21	0.39	0.40	0.00	0.34	0.66
Bayes		0.21	0.39	0.40	0.05	0.31	0.64
MLE		0.03	0.02	0.95	0.03	0.01	0.96
OLS		0.03	0.02	0.95	0.06	0.02	0.92
ICLS	$i = 3$	0.03	0.02	0.95	0.05	0.02	0.93
Bayes		0.03	0.02	0.95	0.05	0.02	0.93

4 In case when the OLS satisfies all the inequality constraints, it is equivalent to the ICLS and very similar to the Bayesian estimates. The purpose of this example is to illustrate the performance of different estimators in case when OLS is not consistent with the imposed constraints. Therefore, the first three simulated series that result with the OLS estimator with all the constraints satisfied were discarded and only the fourth trial that produced *problematic* OLS estimates is reported here.

As expected, given a large number of simulated realisations of the chain, i.e. $n = 10000$, all the estimates resemble the true transition probabilities very well. Moreover, all estimation methods produce probabilities that are equal, up to the second decimal place and regression methods based on the macro data are just as good as a micro-based maximum likelihood estimator. For our purpose, however, more interesting results are produced for n = 100, which is a number of observations to be expected in practice. In contrast to the large sample case, most of the estimates depart more strongly from true probabilities here. The maximum likelihood method, exploiting the micro-data, yields the most precise estimates in general. Regarding the regression estimators, the OLS produced one negative estimate and accordingly, the same probability is estimated as zero using the ICLS. The Bayesian posterior mean estimator produces the positive value here. On the other hand, the OLS and ICLS produce better estimates for p_{22} compared to the Bayesian estimator. Beside this, all the regression methods produce similar estimates for the remaining parameters.

In order to be able to judge the ability of individual estimators to mimic the true matrix of transitions P, an appropriate distance measure between matrices is to be applied. For that purpose, the standard matrix 2-norm can be used. According to this distance criterion, the maximum likelihood estimator of matrix P is the closest to the true matrix, with an estimated distance of 0.23. After that, the Bayesian estimator is second best with the distance of 0.3, and the ICLS and OLS are somewhat worse with distances of 0.34 and 0.44, respectively.

5 Results

5.1 Transition probabilities for Croatian bank loan portfolio

This section presents Bayesian estimates of transition matrix underlying the dynamics of loans extended to domestic firms for the period 1999q2 – 2011q2. The properties of Bayesian estimates are finally compared with those of more conventional estimates obtained via the OLS and ICLS estimators. Only the data on aggregate volumes in every risk category is available. Quarterly series of aggregate loans are classified into four broad risk categories, depending on the possibility of collection, i.e. on the expected future cash flows (CNB 2009, Cerovac and Ivičić 2009):

A – placements for which no evidence of impairment is identified on an individual basis

B – placements for which evidence of partial impairment is identified, i.e. partly recoverable placements

C – placements for which evidence of impairment is identified, equal to their carrying amount, i.e. fully irrecoverable placements (delinquent)

D – cumulative charge-offs on loans. This is defined an absorbing risk category[5]

Debtors' timeliness in meeting their obligations to a credit institution is an important criterion integrated in the above classification scheme and implies a downgrade from A to B if a debtor has overdue liabilities for more than 90 to 365 days, and from B to C if a debtor has overdue liabilities for more than 365 days. For more details on classification of loans see CNB (2009).

Migrations between classes A, B, C and D are modelled as a trajectory of a discrete-time, four-state Markov chain by using the data on proportions of loans in every state, as shown in the previous section. Transition probabilities of a chain, i.e. migration probabilities of the loan portfolio are estimated using the draws produced by the Gibbs sampler under the inequality constrained regression framework.

5 The analysis was also conducted for the three-states chain, omitting the absorbing state. Due to the low proportion of loans in *charge-off* category in our sample, the results were not altered significantly in that case. Another robustness check concerns the stationarity of the underlying Markov chain. There was a major change in the classification methodology in 2003. For that reason, the analysis was also conducted ignoring the period before the break. Again, the main results proved to be robust to the change of sample.

In order to start the sampler, the parameters of the prior distribution and an initial value of the chain need to be specified. Motivated by Rodriguez-Yam et al. (2004), I define the prior parameter μ_0 as the constrained least squares estimator, obtained via quadratic programming.[6] Moreover, the scaling constant of the prior covariance matrix σ_0 should be large enough; here it is 100, in order to reflect the high prior uncertainty about the coefficients. Finally, the parameters of the prior for the noise variance should also reflect a lack of prior information and are set to[7] $\nu = \lambda = 0.01$. After defining the priors, 10000 draws from the posterior conditionals for six regression coefficients and noise variance are produced. The draws for remaining coefficients are calculated directly from equality constraints (6). The Bayesian estimator used here is the standard *posterior mean estimator,* that is, sample average of posterior draws.

5.2 Estimation results

Table 2 compares the estimates of transition probabilities of Croatian banks loans to the corporate sector, obtained by the ordinary least squares, OLS (i.e. by ignoring constraints), the constrained least squares estimator, ICLS (obtained via quadratic programming) and the Bayesian estimator.

Table 2 One-quarter transition probabilities, OLS, ICLS and Bayesian estimates (in percentages)

		j = A	*j = B*	*j = C*	*j = D*
OLS		100.04	0.05	–0.08	–0.01
ICLS	*i = A*	99.74	0.23	0.03	0
Bayes		**99.45**	**0.28**	**0.19**	**0.08**
OLS		–16.04	112.42	5.13	–1.51
ICLS	*i = B*	0.00	96.55	3.45	0
Bayes		**2.87**	**91.7**	**4.43**	**1**
OLS		34.18	–26.41	88.94	3.29
ICLS	*i = C*	10.69	0.00	89.31	0
Bayes		**9.74**	**6.05**	**82.16**	**2.05**
OLS		0	0	0	100
ICLS	*i = D*	0	0	0	100
Bayes		**0**	**0**	**0**	**100**

As expected, the OLS estimates do not satisfy the imposed inequality constraints and some of the estimates are greater than unity or smaller than zero.[8] On the other hand, by their construction, both the ICLS and Bayesian estimates satisfy all the constraints. However, there are some differences between them. For example, the ICLS has tendency to produce estimates close to the boundary. The estimated probabilities thus reveal an unattractive property of the ICLS estimator, in that it often produces uninformative corner solutions. Besides that, the absence of any relevant asymptotic theory for the ICLS makes inference about the reliability of the parameters difficult.

Bayesian estimates, on the other hand, are acceptable and easy to interpret. The A-rated loans mostly stay in the same class in the next quarter as well, with the high probability of 0.99. Dynamics of B and C rated loans has not shown such a persistence during recent years. For example, B-rated loans proceed with downgrading to C with a probability of less than 5%. The upgrading probabilities are rather low, less than 3%. On the other

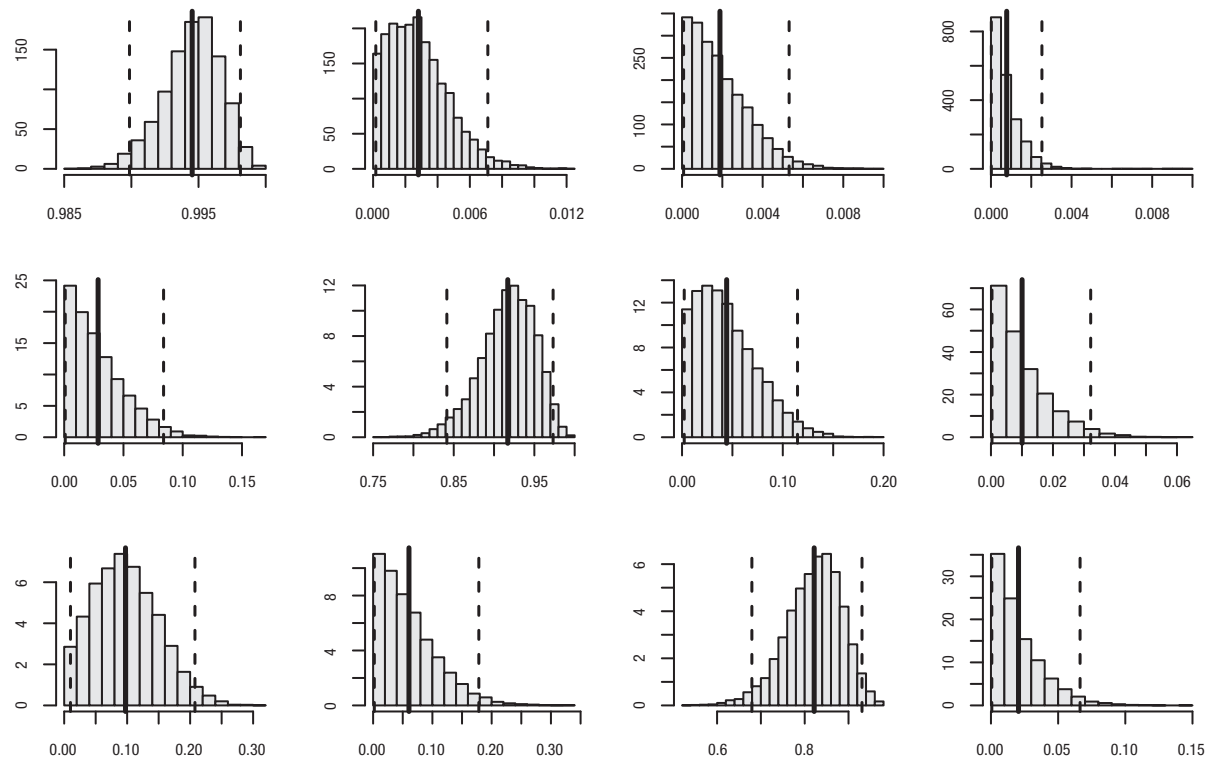
6 See Byrd et al. (1995) and Coleman and Li (1996).

7 It can be shown that the inverse-gamma prior, $IG(\varepsilon,\varepsilon)$ for variance σ^2 converges towards the uniform prior for log $\log\sigma^2$ when $\varepsilon \to 0$. Therefore, setting ε to be small attaches similar densities to all the values from support of the prior which reflects the prior uncertainty about the variance.

8 The rows of transition matrix, however, do sum up to unity by construction.

hand, the constrained estimator predicts zero probability for a B-loan to recover fully and become A again. The loans with the lowest rating C are charged-off with a probability of 2% and the estimated transition probabilities suggest that rating C is not the final (absorbing) state of a typical loan. Again, the constrained estimator predicts zero probability for a C-loan to be charged-off.

Figure 1 Posterior histograms of full conditionals for regression coefficients. The graph in the row i and column j represents the density of $p_{ij} = P(y_t = j | y_{t-1} = i)$. Vertical lines represent the estimates for 0.025 and 0.975 quantiles and posterior mean.



Source: Author's calculation.

Figure 1 shows the estimated densities for conditional posterior distributions of regression coefficients together with the Bayesian 95% credible intervals. In the classical approach, the boundaries of confidence intervals and not the parameters are random quantities. As a consequence, one cannot conclude that there is, say, 0.95 probability for a parameter to lie within the 0.95 confidence interval, given the data. Rather, the confidence interval includes the given value with 0.95 probability, before one has even seen the data. In contrast to that, Bayesian credible intervals allow one to construct intervals such that the probability for a (random) parameter to lie within this interval is 0.95. Estimated histograms illustrate the effect of the truncation and some of the conditionals are characterised by well skewed empirical distributions.

In order to study the speed of convergence of our sampler towards the stationary distribution, it is also useful to compare the running means of posterior draws for transition probabilities with the means calculated over the whole sample (see appendix).

6 Conclusion

In this paper I studied the Gibbs sampler for the parameters of the inequality constrained regression proposed by Rodriguez-Yam et al. (2004). The methodology was applied to the problem of the estimation of the migration probabilities between different rating categories for loans in the portfolios of Croatian banks. The analysis is conducted for loans to the corporate sector. It is assumed that each unit in the portfolio obeys a discrete Markov chain process. In that case, migration probabilities may be estimated under the inequality constrained regression framework, even though only the data on the proportions of loans in every rating class is available. The main findings of the paper may be summarised as follows.

Bayesian estimates based on the Gibbs sampler are acceptable and easy to interpret. The loans with the highest quality mostly stay in the same class in the next quarter as well, with a probability of 0.99. In other words, the A rating has been nearly the absorbing state for a typical loan in Croatia during the past decade. On the other hand, B and C rated loans have not shown such a persistence during the last ten years. For example, B-rated loans proceed with their downgrading to C with a probability of less than 5%. The probabilities of upgrading are rather low, less than 3%. The loans with the lowest rating C are charged-off with a probability of 2% and the estimated transition probabilities suggest that rating C is not the final (absorbing) state of a typical loan.

In contrast to Bayesian estimates, the standardly used least squares appear to be more *problematic*. The ordinary least squares estimates do not satisfy inequality restrictions and, therefore, cannot be interpreted as transition probabilities. On the other hand, inequality constrained least squares by their construction do satisfy all the restrictions, but show a tendency to produce boundary solutions which makes them hard to interpret. For example, the constrained estimator predicts zero probability for a B-loan to recover fully and become A again. Similarly, the constrained estimator predicts zero probability for a C-loan to be charged-off. Although the micro-data on individual loans is not available and, accordingly, this particular transition cannot be observed directly, it is unlikely that such events have not been recorded in the past decade.

This analysis deals with supervisory, aggregate, data on proportions only and, hence, it is difficult to judge how well the obtained estimates resemble the underlying structural transition probabilities. In order to investigate this issue in more detail, an illustrative example based on the simulated paths of the micro-units is considered. If the time dimension is not very large ($n=100$), regression estimates based on the aggregate data may depart significantly from the true probabilities. However, both the Bayesian estimates and the estimates obtained by constrained linear squares approximate the true transitions well. As expected, the maximum likelihood estimates based on the micro data produce the estimates with the highest precision and should always be used when micro data is available. Using the matrix norm as a distance measure, Bayesian estimates appear to be somewhat closer to the true parameters than the inequality constrained estimates. The OLS, again, performs poorly.

Regarding the characteristics of the implemented sampler, the posterior sample has a quickly decaying autocorrelation function (appendix). As a consequence, posterior means converge very fast and simulated draws explore posterior distribution efficiently. The convergence is mostly achieved after only around 1000 iterations, which is considerably faster than with alternative specifications, where tens or hundreds of thousands of iterations are needed. Moreover, the specification easily handles a large number of, potentially collinear, restrictions. This is a property that enables the estimation of transition probabilities of a discrete Markov chain and does not characterise alternative Gibbs sampler for constrained regression (Geweke 1996).

This analysis additionally confirmed all the good properties that characterise the sampler proposed by Rodriguez-Yam et al. (2004). The results obtained here suggest that estimates based on this sampler should be considered as an alternative to standard methods when estimating Markov transition matrices in practice.

References

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995): *A limited memory algorithm for bound constrained optimization*, SIAM J. Scientific Computing, 16, 1190-1208.

Casella, G., and E. I. George (1992): *Explaining the Gibbs sampler*, The American Statistician, 46 (3): 167-174.

Cerovac, S., and L. Ivičić (2009): *Credit Risk Assessment of Corporate Sector in Croatia*, The paper presented on the 15th Dubrovnik economic conference.

Chamberlain, G., and E. E. Leamer (1976): *Matrix weighted averages and posterior bounds*, Journal of the Royal Statistical Society, Series B, 73-84.

Christodoulakis, G. A. (2006): *Markovian Credit Risk Transition Probabilities under Non-Negativity Constraints for the US Portfolio 1984-2004*, Manchester Business School, Working paper series.

CNB (2009): *Decision on the classification of placements and off-balance sheet liabilities of credit institutions*.

Coleman, T. F., and Y. Li (1996): *A Reflective Newton Method for Minimizing a Quadratic Function Subject to Bounds on Some of the Variables*, SIAM Journal on Optimization, Vol. 6, Number 4, pp. 1040-1058.

Davis, W. W. (1978): *Bayesian analysis of the linear model subject to linear inequality constraints*, Journal of the American Statistical Association, 573-79.

Geman, S., and D. Geman (1984): *Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741.

Feller, W. (1971): *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd Edition, Wiley.

Geweke, J. (1986): *Exact Inference in the Inequality Constrained Normal Linear Regression Model*, Journal of Applied Econometrics, John Wiley & Sons, Ltd., Vol. 1(2), pp. 127-141.

Geweke, J. (1991): *Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities*, The paper was prepared for a presentation at the meeting Computing Science and Statistics: the Twenty-Third Symposium on the Interface, Seattle, 22-24 April 1991.

Geweke, J. (1996): *Bayesian Inference for Linear Models Subject to Linear Inequality Constraints*, in Modeling and Prediction: Honouring Seymour Geisser, eds. W. O. Johnson, J. C. Lee, and A. Zellner, New York, Springer, pp. 248-263.

Geyer, C., web page: www.stat.umn.edu/126charlie/.

Hajivassiliou, V. A., and D. McFadden (1990): *The Method of Simulated Scores for the Estimation of LDV Models with an Application to External Debt Crisis*, Cowles Foundation Discussion Papers 967, Cowles Foundation, Yale University.

Hajivassiliou, V. A., D. McFadden, and P. Ruud (1996): *Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results*, Journal of Econometrics, Elsevier, Vol. 72(1-2), pp. 85-134.

Judge, G. G., and T. Takayama (1966): *Inequality Restrictions In Regression Analysis*, Journal of the American Statistical Association, 61, 166-181.

Jones, M. T. (2005): *Estimating Markov Transition Matrices Using Proportions Data: An Application to Credit Risk*, IMF Working Paper WP/05/219.

Kelton, C. M. L. (1981): *Estimation of Time-Independent Markov Processes with Aggregate Data: A Comparison of Techniques*, Econometrica, Vol. 49, No. 2, pp. 517-518.

Kloek, T., and H. K. Van Dijk (1978): *Bayesian estimates of equation system parameters: an application of integration by Monte Carlo*, Econometrica, 46, 1-19.

Lee, T. C., G. G. Judge, and A. Zellner (1968): *Maximum Likelihood and Bayesian Estimation of Transition Probabilities*, Journal of the American Statistical Association, Vol. 63, No. 324, pp. 1162-1179.

Lee, T. C., G. G. Judge, and T. Takayama (1965): *On estimating the transition probabilities of a Markov process*, Journal of Farm Economics, Vol. 47, No. 3, 742-762.

Lovell, M. C., and E. Prescott (1970): *Multiple Regression with inequality constraints: Pretesting Bias, Hypothesis Testing, and Efficiency*, Journal of the American Statistical Association, 65, 913-925.

MacRae, E. C. (1977): *Estimation of Time-Varying Markov Processes with Aggregate Data*, Econometrica, Vol. 45(1), pp. 183-198.

Metropolis, N., and S. Ulam (1949): *The Monte Carlo Method*, Journal of the American Statistical Association, Vol. 44, No. 247 (Sep., 1949), pp. 335-341.

O'Hagan, A. (1973): *Bayes estimation of a convex quadrature*, Biometrika, 60, 565-571.

Robert, C. P., and G. Casella (2004): *Monte Carlo Statistical Methods,* Springer-Verlag, New York.

Rodriguez-Yam, G., R. Davis, and L. Scharf (2004): *Efficient Gibbs sampling of truncated multivariate Normal with application to constrained linear regression,* Working paper, Colorado State University, Fort Collins, CO.

Appendix

Posterior distribution for noise variance

In order to show (21) let us note that the prior for σ^2 implies:

$$f(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\nu+1} e^{-\frac{1}{\lambda\sigma^2}}.$$

Using the Bayes theorem it follows:

$$f(\sigma^2 \,|\, \beta, y) \propto L(\sigma^2 \,|\, \beta, y) f(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2} SS(\beta)} \left(\frac{1}{\sigma^2}\right)^{\nu+1} e^{-\frac{1}{\lambda\sigma^2}} = \left(\frac{1}{\sigma^2}\right)^{\frac{n+2\nu}{2}+1} e^{-\frac{\lambda SS(\beta)+2}{2\lambda\sigma^2}},$$

and comparing this with inverse-gamma density we have:

$$\sigma^2 \Big| (\beta, y) \sim IG\left(\frac{n+2\nu}{2}, \frac{2\lambda}{\lambda SS(\beta)+2}\right),$$

or

$$\sigma^{-2} \Big| (\beta, y) \sim G\left(\frac{n+2\nu}{2}, \frac{2\lambda}{\lambda SS(\beta)+2}\right).$$

Moreover, the characteristic function of gamma density is given by:

$$\varphi_{G\left(\frac{n+2\nu}{2}, \frac{2\lambda}{\lambda SS(\beta)+2}\right)}(t) = \left(1 - \frac{2\lambda}{\lambda SS(\beta)+2} it\right)^{-\frac{n+2\nu}{2}}.$$

Recalling that gamma distribution $G(n/2, 2)$ is equal to $\chi^2(n)$ by definition, it follows immediately:

$$\varphi_{\frac{\chi^2(n+2\nu)}{SS(\beta)+2\lambda^{-1}}}(t) = \varphi_{\chi^2(n+2\nu)}\left(\frac{t}{SS(\beta)+2\lambda^{-1}}\right) = \left(1 - \frac{2}{SS(\beta)+2\lambda^{-1}} it\right)^{-\frac{n+2\nu}{2}} =$$
$$= \left(1 - \frac{2\lambda}{\lambda SS(\beta)+2} it\right)^{-\frac{n+2\nu}{2}} = \varphi_{G\left(\frac{n+2\nu}{2}, \frac{2\lambda}{\lambda SS(\beta)+2}\right)}(t).$$

Finally, using the *Uniqueness Theorem for Characteristic Functions* (Feller 1971), random variables $\frac{\chi^2(n+2\nu)}{SS(\beta)+2\lambda^{-1}}$ and $G\left(\frac{n+2\nu}{2}, \frac{2\lambda}{\lambda SS(\beta)+2}\right)$ are equally distributed and (21) holds. In similar way, (20) can be shown as well.

Properties of the truncated normal distribution

Proposition 1 *Let* $X \sim N_R(\mu, \Sigma)$ *where* $R \in \mathbb{R}^k$ *has positive Lebesque measure and* Σ *the positive definite. Let* $Y := AX$, *where* A *is* $r \times k$ *matrix with* $r \leq k$. *Then* $Y \sim N_T(A\mu, A\Sigma A^\tau)$, *where* $T := \{Ax : x \in R\}$.

Proposition 2 *Let us partition* X, μ, Σ *in the following way*

$$X = \begin{pmatrix} X_1 \\ X_k \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_k \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_1 \\ \Sigma_1^\tau & \sigma_{kk} \end{pmatrix}$$

where X_k is the univariate components of X. Then the univariate conditionals take on the form:

$$X_k \,|\, (X_1 = x_1) \sim N_{R_k}(\mu_k^*, \sigma_{kk}^*),$$

where

$$\mu_k^* = \mu_k + \Sigma_1^\tau \Sigma_{11}^{-1}(x_1 - \mu_1)$$

$$\sigma_{kk}^* = \sigma_{kk} - \Sigma_1^\tau \Sigma_{11}^{-1} \Sigma_1,$$

for new constrained set:

$$R_k := \{x_k \in R : (x_1, x_k) \in R\}.$$

Bayesian statistics

Standardly, the model's parameters are estimated via *maximum likelihood methods* and *method of moments.* Under the classical framework, the parameter is seen as a fixed, albeit unknown variable. The prior beliefs about the parameters to be estimated are irrelevant for the estimation process. On the other hand, Bayesian inference is based on a different idea. Here, the parameter is seen as a random variable with two distributions of interest. First is *the prior distribution* and reflects our knowledge on the parameter even before the data is gathered and, second, *the posterior distribution* that is *conditional on observed data*. The prior and posterior relate to each other via the Bayes theorem as follows:

$$f(\beta|\mathcal{D}) = \text{Bayes theorem} = \frac{f(\mathcal{D}|\beta)f(\beta)}{f(\mathcal{D})} \propto f(\mathcal{D}|\beta)f(\beta), \tag{28}$$

where β denotes (random) parameter of interest, $\mathcal{D}$ is the data, $f(\beta)$ is the prior, $f(\beta|\mathcal{D})$ is the posterior and $f(\mathcal{D}|\beta) = L(\beta|\mathcal{D})$ is the likelihood function of a model. In contrast to standard estimation strategies, under the Bayesian framework, both the prior knowledge about the parameter (the prior) as well as the sample information (the likelihood function) are used to make inferences about the parameter of interest. However, although the possibility of inclusion of the prior beliefs about the parameter into the estimation process is often seen as an important advantage over the standard estimation techniques, it is also the main target of criticism of the Bayesian models. The fact that individuals form various prior views about the parameter of interest, may result in substantially different posterior distribution which raises the question of consistency of the approach.

Often, a Bayesian model is conveniently defined by the posterior distribution belonging to the same family as the prior. A well known example is *the conjugate Normal-Gamma prior* in the normal linear regression model. In other words, if normal prior is assumed for regression coefficients and gamma prior for noise variance, the marginal posteriors for coefficients and the variance are normally and gamma distributed as well. This can be readily seen from (1) and the form of the likelihood function of a Gaussian linear model. In this case the theoretical properties of the posterior are all we need to make an inference about the parameters. Although the whole posterior distribution carries the information about the parameter to be estimated it is a single point estimate that is typically of interest for applications. A usual choice for a point estimate is a simple *posterior mean estimator,* that is, the expectation of posterior distribution.

However, in many cases the posterior is easily calculated up to the constant only and $f(\beta|\mathcal{D}) \propto f(\mathcal{D}|\beta)f(\beta)$ may be the *best guess* about the posterior. The reason for that is typically the fact that the normalising constant is very difficult to derive analytically via high-dimensional integration. Moreover, the posterior need not to belong to any of the standard probability distributions and the exact inference on the posterior becomes very difficult. Traditionally, this has been the main drawback of many Bayesian models and their application. The Markov Chain Monte Carlo methods (MCMC) have proved to be very efficient in providing the solution for the problem by using the simulations.

Markov Chain Monte Carlo methods

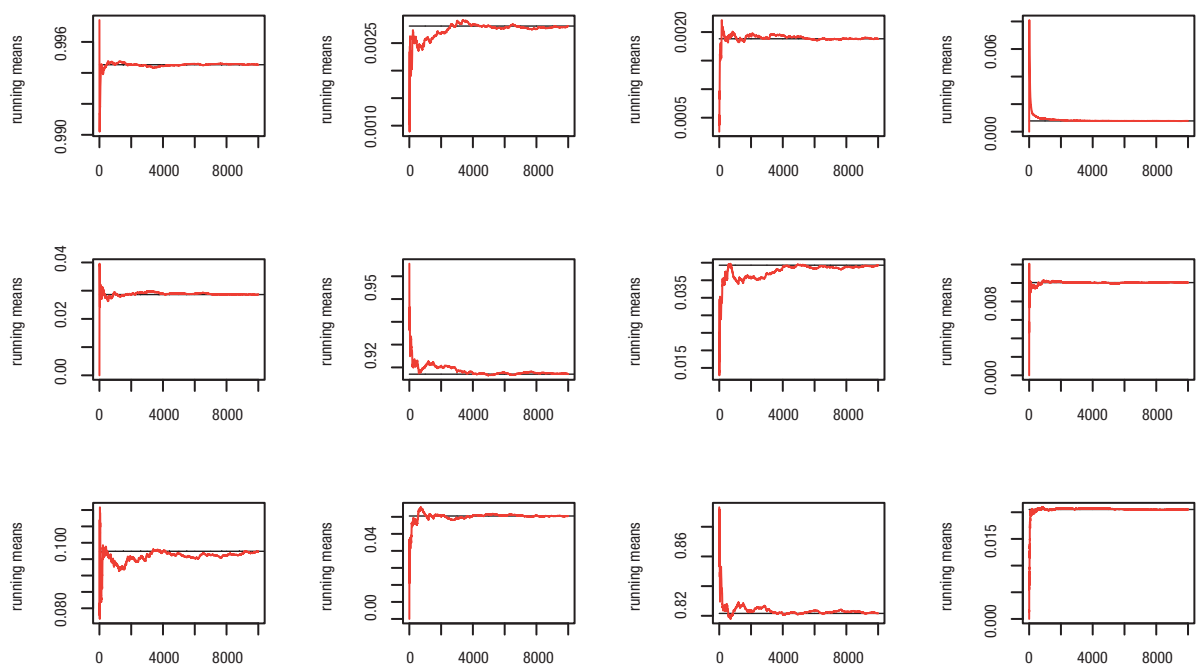
The Markov Chain Monte Carlo methods (MCMC) are designed to draw a random sample from a distribution that may be known up to the normalising constant only. The basic idea underlying the MCMC is to simulate a path of a Markov Chain that has a distribution of interest as its *limiting distribution*. In terms of expression (1) it is sufficient to simulate a chain that *converges* to $f(\beta|\mathcal{D})$. In this way, under relatively mild conditions (see Robert and Casella 2004 for details), a sufficiently long simulated path of a Markov Chain may be seen as a sample from that distribution. Although the exact inference about the posterior might not be convenient, an approximate sample from the distribution of interest may often be drawn via MCMC and all the sample statistics can easily be obtained. It is important to note that the sample produced by MCMC is a trajectory of a Markov Chain and, accordingly, cannot be seen as a series of independent draws from a limiting distribution. Still, an *ergodic Markov chain,* under some additional assumptions, inherits all the *good* properties from the *Law of large numbers* (LLN) and the *Central limit theorem* (CLT) and, hence, all the sample statistics are well behaved.

The MCMC methods originate from the *Metropolis algorithm* (Metropolis and Ulam 1949) as an attempt of a group of physicists to calculate complicated integrals via simulations. Building on this algorithm, the two *flagship* MCMC methods are constructed, *the Metropolis-Hastings* algorithm and, finally, its special case, *the Gibbs sampler* (see Geman and Geman 1984).

Convergence diagnostics and conditional posteriors

The reliability of Bayesian estimators heavily depends on two factors. First, the dispersion of posterior distribution is important as it indicates the precision of the estimates. The smaller the variance, more confident we are about the Bayesian estimates. Second, the Bayesian estimators rely on the Gibbs draws under the

Figure 2 Running means of full conditionals for regression coefficients. For each series the means calculated over the whole sample are represented by a full line.



Source: Author's calculation.

assumption that the underlying chain has indeed converged towards the posterior distribution. The easiest way to check this is by inspecting the graphs of *the running means* over time.

In order to check the speed of convergence of our sampler towards the stationary distribution, Figure 2 compares the running means of posterior draws for transition probabilities with the means calculated over the whole sample. For all the parameters, running means are close to the overall mean, after only few hundreds of iterations. It is apparent how averages depend heavily on the initial state of the chain.[9] Therefore, a so called *burn-in* sample consisted of the starting draws is often discarded from further analysis. Clearly, throwing away some iterations at the beginning of an MCMC run is just another way to initialise a chain, by using the point that is more likely to be drawn from the posterior. However, discarded values are possible outcomes from the posterior as well, but not in the very short run. Instead of simply throwing away the initial observation of a chain, in this paper, somewhat longer MCMC sequences are considered and initial realisations do not affect sample statistics of the posterior draws. Initial observations are, therefore, not discarded from the analysis.[10]

9 In order to check the sensitivity of the posterior sample statistics to the specification of the initial states, the sampler was started from different points from the support of the posterior. Sample statistics would converge quickly for different initial points.

10 In an interesting discussion Charles Geyer (Geyer's web page) explains why and when *burn-in* is unnecessary. *Burn-in is only one method, and not a particularly good method, of finding a good starting point. The name burn-in comes from electronics. Many electronics components fail quickly. Those that don't are a more reliable subset. So a burn-in is done at the factory to eliminate the worst ones. Markov chains don't work the same way. Markov chain failure (nonconvergence) is different from electronic component failure. Running longer may cure the first, but a dead transistor is dead forever. Thus burn-in is a bad term in MCMC, but there's more wrong than just the word, there's something fishy about the whole concept...*

The following Working Papers have been published:

No.	Date	Title	Author(s)
W-1	December 1999	Croatia in the Second Stage of Transition, 1994–1999	Velimir Šonje and Boris Vujčić
W-2	January 2000	Is Unofficial Economy a Source of Corruption?	Michael Faulend and Vedran Šošić
W-3	September 2000	Measuring the Similarities of Economic Developments in Central Europe: A Correlation between the Business Cycles of Germany, Hungary, the Czech Republic and Croatia	Velimir Šonje and Igeta Vrbanc
W-4	September 2000	Exchange Rate and Output in the Aftermath of the Great Depression and During the Transition Period in Central Europe	Velimir Šonje
W-5	September 2000	The Monthly Transaction Money Demand in Croatia	Ante Babić
W-6	August 2001	General Equilibrium Analysis of Croatia's Accession to the World Trade Organization	Jasminka Šohinger, Davor Galinec and Glenn W. Harrison
W-7	February 2002	Efficiency of Banks in Croatia: A DEA Approach	Igor Jemrić and Boris Vujčić
W-8	July 2002	A Comparison of Two Econometric Models (OLS and SUR) for Forecasting Croatian Tourism Arrivals	Tihomir Stučka
W-9	November 2002	Privatization, Foreign Bank Entry and Bank Efficiency in Croatia: A Fourier-Flexible Function Stochastic Cost Frontier Analysis	Evan Kraft, Richard Hofler and James Payne
W-10	December 2002	Foreign Banks in Croatia: Another Look	Evan Kraft
W-11	October 2003	The Impact of Exchange Rate Changes on the Trade Balance in Croatia	Tihomir Stučka
W-12	August 2004	Currency Crisis: Theory and Practice with Application to Croatia	Ivo Krznar
W-13	June 2005	Price Level Convergence: Croatia, Transition Countries and the EU	Danijel Nestić
W-14	March 2006	How Competitive Is Croatia's Banking System?	Evan Kraft
W-15	November 2006	Microstructure of Foreign Exchange Market in Croatia	Tomislav Galac, Ante Burić, Ivan Huljak
W-16	December 2006	Short-Term Forecasting of Inflation in Croatia with Seasonal ARIMA Processes	Andreja Pufnik and Davor Kunovac
W-17	February 2008	Modelling of Currency outside Banks in Croatia	Maroje Lang, Davor Kunovac, Silvio Basač and Željka Štaudinger
W-18	June 2008	International Business Cycles with Frictions in Goods and Factors Markets	Ivo Krznar
W-19	December 2008	Use of the Hedonic Method to Calculate an Index of Real Estate Prices in Croatia	Davor Kunovac, Enes Đozović, Gorana Lukinić, Andreja Pufnik
W-20	May 2009	Contagion Risk in the Croatian Banking System	Marko Krznar
W-21	October 2009	Optimal International Reserves of the CNB with Endogenous Probability of Crisis	Ana Maria Čeh and Ivo Krznar
W-22	December 2009	The Impact of the Financial Crisis and Policy Responses in Croatia	Nikola Bokan, Lovorka Grgurić, Ivo Krznar, Maroje Lang
W-23	April 2010	Habit Persistence and International Comovements	Alexandre Dmitriev and Ivo Krznar
W-24	April 2010	Capital Inflows and Efficiency of Sterilisation – Estimation of Sterilisation and Offset Coefficients	Igor Ljubaj, Ana Martinis and Marko Mrkalj
W-25	April 2010	Income and Price Elasticities of Croatian Trade – A Panel Data Approach	Vida Bobić
W-26	December 2010	Impact of External Shocks on Domestic Inflation and GDP	Ivo Krznar and Davor Kunovac
W-27	December 2010	The Central Bank as Crisis-Manager in Croatia – A Counterfactual Analysis	Tomislav Galac
W-28	January 2011	A Credit Market Disequilibrium Model And Periods of Credit Crunch	Ana Maria Čeh, Mirna Dumičić and Ivo Krznar
W-29	November 2011	Identifying Recession and Expansion Periods in Croatia	Ivo Krznar

Guidelines to authors

In its periodical publications *Working Papers, Surveys* and *Technical Papers*, the Croatian National Bank publishes scientific and scholarly papers of the Bank's employees and other associate contributors.

After the submission, the manuscripts shall be subject to peer review and classification by the Manuscript Review and Classification Committee. The authors shall be informed of the acceptance or rejection of their manuscript for publication within two months following the manuscript submission.

Manuscripts are submitted and published in Croatian and/or English language.

Manuscripts submitted for publication should meet the following requirements:

Manuscripts should be submitted via e-mail or optical storage media (CD, DVD), accompanied by one printed paper copy. The acceptable text format is Word.

The first page of the manuscript should contain the article title, first and last name of the author and his/her academic degree, name of the institution with which the author is associated, author's co-workers, and the complete mailing address of the corresponding author to whom a copy of the manuscript with requests for corrections shall be sent.

Additional information, such as acknowledgments, should be incorporated in the text at the end of the introductory section.

The second page should contain the abstract and the key words. The abstract is required to be explicit, descriptive, written in third person, consisting of not more than 250 words (maximum 1500 characters). The abstract should be followed by maximum 5 key words.

A single line spacing and A4 paper size should be used. The text must not be formatted, apart from applying bold and italic script to certain parts of the text. Titles must be numerated and separated from the text by double-line spacing, without formatting.

Tables, figures and charts that are a constituent part of the paper must be well laid out, containing: number, title, units of measurement, legend, data source, and footnotes. The footnotes referring to tables, figures and charts should be indicated by lower-case letters (a,b,c…) placed right below. When the tables, figures and charts are subsequently submitted, it is necessary to mark the places in the text where they should be inserted. They should be numbered in the same sequence as in the text and should be referred to in accordance with that numeration. If the tables and charts were previously inserted in the text from other programs, these databases in the Excel format should also be submitted (charts must contain the corresponding data series).

The preferred formats for illustrations are EPS or TIFF with explanations in 8 point Helvetica (Ariel, Swiss). The scanned illustration must have 300 dpi resolution for grey scale and full colour illustration, and 600 dpi for lineart (line drawings, diagrams, charts).

Formulae must be legible. Indices and superscript must be explicable. The symbols' meaning must be given following the equation where they are used for the first time. The equations in the text referred to by the author should be marked by a serial number in brackets closer to the right margin.

Notes at the foot of the page (footnotes) should by indicated by Arabic numerals in superscript. They should be brief and written in a smaller font than the rest of the text.

References cited in the text are listed at the last page of the manuscript in the alphabetical order, according to the authors' last names. References should also include data on the publisher, city and year of publishing.

Publishing Department maintains the right to send back for the author's revision the accepted manuscript and illustrations that do not meet the above stated requirements.

All contributors who wish to publish their papers are welcome to do so by addressing them to the Publishing Department, following the above stated guidelines.

The Croatian National Bank publications:

Croatian National Bank – Annual Report
Regular annual publication surveying annual monetary and general economic developments as well as statistical data.

Croatian National Bank – Semi-annual Report
Regular semi-annual publication surveying semi-annual monetary and general economic developments and statistical data.

Croatian National Bank – Quarterly Report
Regular quarterly publication surveying quarterly monetary and general economic developments.

Banks Bulletin
Publication providing survey of data on banks.

Croatian National Bank – Bulletin
Regular monthly publication surveying monthly monetary and general economic developments and monetary statistics.

Croatian National Bank – Working Papers
Occasional publication containing shorter scientific papers written by the CNB employees and associate contributors.

Croatian National Bank – Surveys
Occasional publication containing scholarly papers written by the CNB employees and associate contributors.

Croatian National Bank – Technical Papers
Occasional publication containing papers of informative character written by CNB employees and associate contributors.

The Croatian National Bank also issues other publications such as, for example, numismatic issues, brochures, publications in other media (CD-ROM, DVD), books, monographs and papers of special interest to the CNB as well as proceedings of conferences organised or co-organised by the CNB, educational materials and other similar publications.

ISSN 1331-8586 (print) • ISSN 1334-0131 (online)