Bayesian Inference in Dynamic Disequilibrium Models : an Application to the Polish Credit Market

L. Bauwens and M. Lubrano

Discussion Paper 2006-27

Département des Sciences Économiques
de l'Université catholique de Louvain

CORE DISCUSSION PAPER
2006/50

BAYESIAN INFERENCE IN DYNAMIC DISEQUILIBRIUM MODELS: AN APPLICATION TO THE POLISH CREDIT MARKET

Luc Bauwens[1] and Michel Lubrano[2]

23 May 2006

Abstract

We review Bayesian inference for dynamic latent variable models using the data augmentation principle. We detail the difficulties of simulating dynamic latent variables in a Gibbs sampler. We propose an alternative specification of the dynamic disequilibrium model which leads to a simple simulation procedure and renders Bayesian inference fully operational. Identification issues are discussed. We conduct a specification search using the posterior deviance criterion of Spiegelhalter, Best, Carlin, and van der Linde (2002) for a disequilibrium model of the Polish credit market.

Keywords: Latent variables, disequilibrium models, Bayesian inference, Gibbs sampler, credit rationing.
JEL Classification: C11, C32, C34, E51.

[1]CORE and Department of Economics, Université catholique de Louvain.
[2]GREQAM and CNRS, Marseille.

The paper is a revised version of the paper written during the sabbatical stay of Luc Bauwens in 2004 at GREQAM, whose support is gratefully acknowledged. We would like to thank deeply Michal Gradzewicz and Andrzej Kociecki from the National Bank of Poland for providing us with the data and the necessary explanations. The paper benefited greatly from comments by seminar participants at Greqam and by the participants of the Thirty First International Conference MACROMODELS 2004 held in Belchatow-Poland December 1-4, 2004 where this paper was presented as an invited conference, and from the comments of two anonymous referees.

This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

This paper discusses the Bayesian treatment of a class of regression models involving latent variables with a possible dynamic specification. These models would be rather simple to analyze if the latent variables were observed. However, the structural equations are completed by an observation rule implying that only a censoring function of the latent variables is observable. Typical examples are Probit and Tobit regression models, and various formulations of disequilibrium models. Classical inference in these models is traditionally a difficult task because the likelihood function involves the computation of an integral. In the static case, the integral is of dimension one for each observation so that the maximum likelihood approach is not difficult to implement. We can quote Tobin (1958) for the static Tobit model and Maddala and Nelson (1974) for static disequilibrium models. However, the dynamic case increases tremendously the dimension of the integral so that the problem becomes rapidly untractable as shown in Dagenais (1982) for the Tobit model with autocorrelated errors at the order one. One has to resort to simulated maximum likelihood. Lee (1999) solves the case of the dynamic Tobit model and Laroque and Salanie (1993) and Lee (1997) tackle the case of dynamic disequilibrium models. However, these methods, based on simulations, are relatively computer intensive and involve non-trivial algorithms. Moreover, Laroque and Salanie (1993) report that it may be difficult to compute accurate standard errors of the estimates.

The data augmentation principle of Tanner and Wong (1987) was introduced to deal with missing value problems. First seen as a stochastic EM algorithm, it was soon reinterpreted as a Gibbs sampler and used as such for Bayesian inference in the static Tobit model by Chib (1992) and by Zangari and Tsurumi (1996) and Wei (1999) for the dynamic Tobit model. Manrique and Shephard (1998) investigate the use of simulation methods for likelihood and Bayesian inference in dynamic latent variable models and favour the use of the scan sampler algorithm of de Jong (1997) based on the Kalman filter. They make suggestions for Bayesian inference in dynamic disequilibrium models.

The purpose of this paper is first to review the data augmentation principle when applied to latent variable models. We then detail the difficulties raised by the dynamic case on three examples and especially for the dynamic disequilibrium model. This is the object of Section 2. We propose to use an alternative dynamic specification for the disequilibrium model which is originally due to Ginsburgh, Tishler, and Zang (1980). Bayesian inference becomes quite simple (Section 3) and opens the way to model comparison and model checking (Section 4). We show the applicability of these methods on the rationed credit market in Poland (Section 5). Conclusions are drawn in the last section.

2 Data augmentation for latent variable models

Let us consider a multivariate latent variable $\tilde{y}_t$ of $\mathbb{R}^p$ generated by the linear multivariate regression model

$$A(L)\tilde{y}_t = B(L)x_t + u_t, \tag{1}$$

where u_t is a Gaussian white noise of $\mathbb{R}^p$ with zero mean and variance-covariance matrix Σ, and x_t a set of observed exogenous variables of $\mathbb{R}^k$. $A(L)$ and $B(L)$ are matrices of lag polynomials defined as

$$A(L) = I_p - A_1 L - A_2 L - \cdots A_r L^r \qquad B(L) = B_0 + B_1 L + \cdots + B_s L^s. \tag{2}$$

where the A_i and B_i are square matrices of parameters of dimension p. We denote by θ the vector containing all the unrestricted parameters contained in $A(L)$, $B(L)$ and Σ. If $\tilde{y}_t$ were observed, this model would be a classical multivariate dynamic regression model. The difficulty comes from the fact that the latent variable $\tilde{y}_t$ is only partially observed by means of a deterministic censoring rule $g(.)$ which relates the observed variable y_t to the unobserved $\tilde{y}_t$. Being a censoring rule, this is not a one-to-one function. We start by briefly discussing three examples to clear up ideas.

Example 1: *Tobit models.* The dynamic univariate Tobit model is obtained with $p = 1$ in (1) together with the following observation rule:

$$y_t = \max(0, \tilde{y}_t). \tag{3}$$

The latent variable is observed only when it is positive, otherwise zero is observed. The censoring rule is thus perfectly deterministic and we know exactly which observations are censored.

Example 2: *Disequilibrium models.* A dynamic version of the disequilibrium model of Maddala and Nelson (1974) is a two equation model describing notional demand d_t and notional supply s_t. It is obtained for $p = 2$ in (1) while noting $\tilde{y}_{1t} = d_t$ and $\tilde{y}_{2t} = s_t$. As prices are supposed to be sticky, adjustment has to be made by quantities. The observed exchanged quantity $q_t = y_t$ is given by the minimum of demand and supply:

$$q_t = \min(d_t, s_t). \tag{4}$$

This censoring rule is still deterministic, but this time a new difficulty is introduced because we do not know which of demand or supply is observed. We can only compute a probability that an observation is allocated to the demand or supply regime.

Example 3: *Stochastic volatility.* This model does not fall in the class of censoring models, but it gives an interesting illustration of the problems raised by a dynamic latent variable. A univariate latent variable, noted in this particular example h_t, follows the autoregressive process of order one, without exogenous variables,

$$h_t = (1 - \rho)\mu + \rho h_{t-1} + u_t \tag{5}$$

where $u_t \sim N(0, \sigma^2)$. There is no observation rule for h_t like in the previous examples, but h_t is related to an observed variable y_t through its variance by assuming that

$$y_t | h_t \sim N(0, \exp(h_t)). \tag{6}$$

Thus, this assumption can be interpreted as an indirect observation rule for h_t. This type of model is used for financial returns.

The latent variable model (1) can be treated in a unified framework for Bayesian inference by data augmentation. The basic idea is quite simple. If $\tilde{y}$ were observed, (1) would be a Gaussian regression model. The posterior density of θ would belong to the natural conjugate class under a suitable prior distribution, providing thus analytical results, even for the stochastic volatility model (see e.g. Robert and Casella (1999), page 435). The idea of data augmentation initiated by Tanner and Wong (1987) consists in simulating the latent variable conditionally on a given value of θ and on the observations, and then iterating using a Gibbs sampler. Under relatively mild conditions, the chain converges to draws of the joint posterior density of both $\tilde{y}$ and θ. Conditionally on y and $\tilde{y}$, it is relatively easy to get draws of θ from its conditional posterior density. The simulation of the distribution of $\tilde{y}$ conditionally on θ and y is however more demanding because one has to take into account firstly the censoring rule (but this can be simple) and secondly the dynamic behaviour of $\tilde{y}$ which raises special difficulties. It is thus useful to distinguish between the static and the dynamic cases for expository purposes.

2.1 A Gibbs sampler for the static case

Likelihood inference can be difficult in latent variable models because the likelihood function involves an integral. In general, we can write

$$L(\theta; y) = \int L(\theta; \tilde{y}, y) d\tilde{y}, \tag{7}$$

where $\tilde{y} = [\tilde{y}_1, \ldots, \tilde{y}_T]$ and $y = [y_1, \ldots, y_T]$. In static models, the above integral of dimension T factorizes into a product of T integrals of dimension one, but this is not the case in dynamic models. The data augmentation principle introduced in Tanner and Wong (1987) avoids to compute such an integral by considering the conditional posterior density of θ, conditional on both the observed sample and the latent process, so that

$$\varphi(\theta|\tilde{y}, y) \propto L(\theta; \tilde{y}, y)\, \varphi(\theta), \tag{8}$$

where $\varphi(\theta)$ is the prior density. The problem contained in (8) is of course that $\tilde{y}$ is not observed. The data augmentation principle consists in simulating $\tilde{y}$ within a Gibbs sampler to finally simulate sequentially (in the static case)

$$\begin{aligned} \tilde{y}^{j+1} &\sim f(\tilde{y}|\theta^j, y), \\ \theta^{j+1} &\sim \varphi(\theta|\tilde{y}^{j+1}, y), \end{aligned} \tag{9}$$

where the superscript j indicates the iteration number of the Gibbs sampler. In general, the conditional posterior density $\varphi(\theta|\tilde{y}, y)$ does not depend on y because y is a deterministic function of $\tilde{y}$ due to the deterministic observation rule. Given the assumptions on the latent model, $\varphi(\theta|\tilde{y})$ easily belongs to a well known class of density functions such as the normal inverted gamma density if the prior is chosen in an appropriate conjugate class.

Example 1: *(continued) Static Tobit models.* This is the most simple case and its solution is due to Chib (1992). Let us write the model as

$$\begin{aligned} \tilde{y}_t &= x_t'\beta + u_t \\ y_t &= \max(0, \tilde{y}_t) \end{aligned} \tag{10}$$

with $u_t \sim \mathrm{N}(0, \sigma^2)$. The conditional posterior density of β and σ^2 is a Normal inverted-gamma-2 if, for instance, the diffuse prior $\varphi(\beta, \sigma^2) \propto 1/\sigma^2$ is used. It is convenient to reorder the observations so as to define a T_1-vector y_u of unobserved components corresponding to the censored observations and a T_2-vector y_o of uncensored observations in order to build $y^* = [y_u', y_o']'$. Let us call X the $T \times k$ matrix of corresponding exogenous variables. The conditional posterior densities of β and σ^2 are then

$$\begin{aligned} \varphi(\beta|\sigma^2, y^*) &= f_N(\beta|\hat{\beta}, \sigma^2(X'X)^{-1}) \\ \varphi(\sigma^2|y^*) &= f_{I\gamma 2}(\sigma^2|s, T), \end{aligned} \tag{11}$$

where f_N and $f_{I\gamma 2}$ denote respectively the normal and the inverted-gamma2 densities, $\hat{\beta} = (X'X)^{-1}X'y^*$, and $s = y^{*'}y^* - y^{*'}X(X'X)^{-1}X'y^*$.

Since positive draws must be avoided, the conditional distribution of y_u is the truncated normal (more precisely a product of univariate truncated normal densities) denoted by

$$f(y_u|\beta, \sigma^2) = f_{TN}(y_u|X_u\beta, \sigma^2 I_{T_1}) \tag{12}$$

where X_u is the $T_1 \times k$ matrix of exogenous variables corresponding to the censored observations. A draw of y_u serves to build the vector y^* which becomes the vector of the endogenous variables where the zeros are replaced by the simulated values.

The Gibbs sampler is thus fully defined by three steps:

1. *Simulate y_u^{j+1} according to (12) and construct y^{*j+1}.*
2. *Simulate θ^{j+1} according to (11) using y^{*j+1} to compute conditional posterior moments.*
3. *Go to step 1.*

Example 2: *(continued) Static disequilibrium models.* The Tobit model is simple because we know which are the censored observations. In a disequilibrium model, we can only compute the probability that an observation belongs to the demand or supply regime. Let us consider the following simple model:

$$\begin{aligned} d_t &= x_{1t}'\beta_1 + u_{1t}, \\ s_t &= x_{2t}'\beta_2 + u_{2t}, \\ q_t &= \min(d_t, s_t), \end{aligned} \tag{13}$$

where x_{1t} and x_{2t} are exogenous variables of dimensions k_1 and k_2. The error terms u_{1t} and u_{2t} are assumed IID, jointly $\mathrm{N}(0, \Sigma)$ with Σ being diagonal. We have to simulate d_t when the supply regime is operating and s_t when the demand regime is operating. The probability of being in the demand regime is given by

$$\lambda_t = \Pr(d_t - s_t < 0|\theta) = \Phi\left(\frac{x'_{1t}\beta_1 - x'_{2t}\beta_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right), \tag{14}$$

where Φ is the standard normal cdf. Given a draw θ^j, we evaluate (14) for each observation t and draw a uniform random number v. We have to construct two vectors of dimension T, yd and ys which contain alternatively observations and simulations of the demand and supply. We allocate an observation q_t to the demand vector yd if $v < \Pr(d_t - s_t < 0|\theta^j)$ and generate s_t as a truncated normal

$$s_t \sim TN_{s_t > d_t}(x'_{2t}\beta_2^j, \sigma_2^{2j}) \tag{15}$$

which is then allocated to the supply vector ys. If $v > \Pr(d_t - s_t < 0|\theta^j)$, we allocate q_t to the supply regime and generate d_t for the demand regime as

$$d_t \sim TN_{d_t > s_t}(x'_{1t}\beta_1^j, \sigma_1^{2j}) \tag{16}$$

Given yd and ys, it is then straightforward to find the conditional posterior density of β_1, σ_1^2 and of β_2, σ_2^2 to obtain the desired Gibbs sampler whose iteration j is defined by six steps:

1. *Set $\theta = \theta^{j-1}$ (the draw of the previous iteration).*
2. *Compute the regime probability λ_t using (14) and draw a uniform random number v.*
3. *If $v < \lambda_t$, allocate q_t to yd, draw s_t^j using (15) and allocate it to ys.*
4. *If $v > \lambda_t$, allocate q_t to ys, draw d_t^j using (16) and allocate it to yd.*
5. *Repeat steps 2 to 4 for all observations.*
6. *Draw θ^j from its conditional posterior distribution.*

2.2 The pitfalls of the dynamic case

The difficulties entailed by the dynamic case come solely from the simulation of the latent variable. When the process of the latent variable is dynamic, it is not correct to sample this variable from its conditional distribution (given the past) defined directly by the model. There exists a complex dynamic interaction between the latent and observed variables. In the static case, $f(\tilde{y}_t|y_t, \theta) = f(\tilde{y}_t|\theta)$, but the dynamics of the latent process influences the censoring mechanism and consequently we have to simulate the latent variable from

$$f(\tilde{y}_t|\tilde{y}_{\setminus t}, y, \theta),$$

where $\tilde{y}_{\setminus t}$ denotes all the past and future of $\tilde{y}_t$.

Example 1: *(continued) Dynamic Tobit models.* The dynamic case was treated by Wei (1999). Supposed that we are in an AR(1) process for simplicity. Wei (1999) draws the latent variables by groups of consecutive occurrences which are preceded and followed by one uncensored observation. The conditional distribution of such a group is a truncated multivariate normal. Its mean and variance have to be built recursively.

Example 2: *(continued) Dynamic disequilibrium models.* The procedure advocated by Manrique and Shephard (1998) is exactly the same as in the static case in the sense that the conditional density of $s_t|s_{\backslash t}, \theta$ is still a truncated normal. But this time its mean and variance have to be determined by the scan sampler of de Jong (1997).

Example 3: *(continued) Stochastic volatility.* Bayesian inference was first treated by Jacquier, Polson, and Rossi (1994). The distribution of h_t given its past and the past of y_t is Gaussian while we need in fact the distribution of h_t conditional to y_t and $h_{\backslash t}$. It can be shown that this distribution is given by

$$f(h_t|h_{\backslash t}, y, \theta) \propto f(h_t|h_{\backslash t}, \theta)\, f(y_t|h_t, \theta) \tag{17}$$

where

$$f(h_t|h_{\backslash t}, \theta) = f(h_t|h_{t-1}, h_{t+1}, \theta) = f_N(h_t|\mu_t, \sigma^2/(1+\rho^2)), \tag{18}$$

the conditional mean μ_t being given by

$$\mu_t = \mu + \frac{\rho(h_{t-1} + h_{t+1} - 2\mu)}{1+\rho^2}. \tag{19}$$

It is relatively difficult to sample from (17). Several algorithms have been proposed in the literature and are reviewed in Kim, Shephard, and Chib (1998). Some of them are based on rejection techniques, other on the Metropolis algorithm with a gamma candidate function.

Our goal is to provide a simple, quick and feasible algorithm for Bayesian inference in dynamic disequilibrium models. Empirical work usually involves a specification search that requires to estimate several versions of a model. We have the feeling that this goal cannot be easily reached with the above dynamic version of the Maddala-Nelson specification. Consequently, we propose in the next section to use an alternative specification for which Bayesian inference proves to be much simpler and which has received a great attention in the classical empirical literature.

3 The GTZ specification of disequilibrium models

Ginsburgh, Tishler, and Zang (1980) propose a formulation of the canonical disequilibrium model which led to important empirical work, see for instance the papers collected in Drèze *et al.* (1991). Agents make plans which are modelled by $x'_{1t}\beta_1$ and $x'_{2t}\beta_2$. They realize that theirs plans are incompatible, so that the exchanged quantity is the minimum of their plans plus an error term which is interpreted as an error of observation. In the dynamic case, agents incorporate the past state of the market q_{t-1} to form theirs plans. Using the same notations as before, we can express the model as

$$\begin{aligned} d_t &= \mathrm{E}(\rho_1 q_{t-1} + x'_{1t}\beta_1), \\ s_t &= \mathrm{E}(\rho_2 q_{t-1} + x'_{2t}\beta_2), \\ q_t &= \min(d_t, s_t) + \epsilon_t, \end{aligned} \tag{20}$$

where ϵ_t is a normal error term of zero mean and variance σ^2. Classical inference for this model is much more simple than for the Maddala-Nelson model, since the likelihood function does not involve an integral for latent variables despite of the censoring rule:

$$L(q;\theta) \propto \sigma^{-T} \exp\left(-\frac{1}{2\sigma^2}\sum_t (q_t - \min[\rho_1 q_{t-1} + x'_{1t}\beta_1,\ \rho_2 q_{t-1} + x'_{2t}\beta_2])^2\right). \tag{21}$$

This function can be maximized using an appropriate algorithm that smooths its local discontinuities. As the dynamics relies solely on past observations and not on past latent variables, there are no differences either for inference or for simulation between the static and the dynamic cases.[1]

Richard (1980) and Sneessens (1985) argue that this formulation should be preferred to the more traditional one of Maddala and Nelson (1974), because it displays nicer statistical and empirical properties.

3.1 A more general specification

The above model, especially when considering its likelihood function (21), can be considered as a non-linear regression without latent variables. We can show that it may fully enter the class of censored latent variable models and that the data augmentation principle can be used for Bayesian inference.

Let us introduce the extra latent variable δ_t which can take only two values 0 or 1 depending on the operating regime at time t, and two independent Gaussian error terms $u_{1t} \sim N(0, \sigma_1^2)$ and $u_{2t} \sim N(0, \sigma_2^2)$. We write the model as

$$\begin{aligned} q_t &= \delta_t(\rho_1 q_{t-1} + x'_{1t}\beta_1 + u_{1t}) + (1-\delta_t)(\rho_2 q_{t-1} + x'_{2t}\beta_2 + u_{2t}) \\ \delta_t &= \begin{cases} 1 \quad \text{if} \quad \rho_1 q_{t-1} + x'_{1t}\beta_1 < \rho_2 q_{t-1} + x'_{2t}\beta_2 \\ \\ 0 \qquad\qquad\qquad \text{otherwise.} \end{cases} \end{aligned} \tag{22}$$

The variable variable δ_t is introduced only for notational convenience. It is fully determined once we know the regression parameters. Consequently, it is a redundant parametrisation. Formulation (22) becomes observationally equivalent to the original GTZ specification when $\sigma_1^2 = \sigma_2^2 = \sigma^2$ as we can define $\epsilon_t = \delta_t u_{1t} + (1-\delta_t)u_{2t}$. It is convenient to allow for a different variance between the two regimes.

[1] Maddala (1987) proposes a similar alternative formulation for the dynamic Tobit model where the lagged variable is observed and not latent. Inference is then as simple as in the static case.

Let us redefine the demand (d_t) and supply (s_t) variables of (22), so that they become fully latent variables that can be decomposed between a systematic and a random part:

$$\begin{aligned} d_t &= \rho_1 q_{t-1} + x'_{1t}\beta_1 + u_{1t} = \mu_{dt} + u_{1t} \\ s_t &= \rho_2 q_{t-1} + x'_{2t}\beta_2 + u_{2t} = \mu_{st} + u_{2t}. \end{aligned} \tag{23}$$

We can rewrite the censoring rule as

$$q_t = \min(\mu_{dt}, \mu_{st}) + \delta_t u_{1t} + (1 - \delta_t) u_{2t}. \tag{24}$$

The two latent variables d_t and s_t have a well defined conditional distribution which is a truncated normal with

$$\begin{aligned} d_t &\sim TN_{\mu_{dt} > \mu_{st}}(d_t | \mu_{dt}, \sigma_1^2) \\ s_t &\sim TN_{\mu_{dt} < \mu_{st}}(s_t | \mu_{st}, \sigma_2^2) \end{aligned} \tag{25}$$

to be compared with formulae (15) and (16) of the static Maddala-Nelson specification. In both cases, we have truncated normals. But as the switching rule operates here, not on random variables as in the Maddala-Nelson specification, but on the non-stochastic quantities μ_{st} and μ_{dt}, the truncation is automatically verified once θ is known. Conditionally on θ, we know perfectly to which regime each observation belongs. This has two consequences. Firstly, the conditional regime probability at t is given by a Dirac function $\mathbb{I}(.)$

$$\Pr(\mu_{dt} < \mu_{st} | \theta) = \mathbb{I}(\mu_{dt} < \mu_{st} | \theta). \tag{26}$$

and no longer by the Gaussian cdf. It is either zero or one. However, in the Bayesian framework, where parameters are random, the unconditional regime probabilities take values in the whole range $[0, 1]$ and are defined by

$$\Pr(\mu_{dt} < \mu_{st}) = \int \mathbb{I}(\mu_{dt} < \mu_{st} | \theta) \varphi(\theta | y) d\theta. \tag{27}$$

In a Monte Carlo setting they can be approximated by

$$\Pr(\mu_{dt} < \mu_{st}) \simeq \frac{1}{N} \sum_j \mathbb{I}(\mu_{dt} < \mu_{st} | \theta^j),$$

where θ^j denotes a draw of $\varphi(\theta | y)$.

Secondly, as the regime selection is made on expectations and not on realisations, we observe the demand or supply plans plus an error term. If the error is large and the difference between the plans small, the observed quantity may be larger than or equal to the maximum of the two plans and not to their minimum.

3.2 Bayesian inference

Let us denote by θ the parameters of the model. If we suppose that the covariance between u_{1t} and u_{2t} is zero, we can split θ in two parts θ_1 and θ_2 corresponding to the parameters of each regime. The prior density can be defined independently on θ_1 and θ_2, for example a diffuse prior given by

$$\varphi(\theta_1) \propto \sigma_1^{-2} \qquad \varphi(\theta_2) \propto \sigma_2^{-2}. \tag{28}$$

An informative natural conjugate prior would be a conditional normal density on $\rho_j, \beta_j|\sigma_j^2$ and an inverted-gamma-2 on σ_j^2, for $j = 1, 2$.

Despite the fact that the likelihood function (21) does not involve an integral, it is useful and computationally efficient to devise a Gibbs sampler similar to (9). Let us define two matrices Z_1 and Z_2 corresponding to the $T-1$ observations of both the predetermined variables q_{t-1} and the exogenous variables x_{jt} in the two regime equations and let us denote by γ_j ($j = 1, 2$) the corresponding regression coefficients, each of dimension k_j.

For a given value of γ_1 and γ_2, we know perfectly which regime is operating. The allocation step designed for the static Maddala Nelson specification of example 2 is much simplified. For each t, if $\mu_{dt} < \mu_{st}$, we allocate q_t to yd_t, draw s_t according to

$$s_t \sim N(s_t|\mu_{st}, \sigma_2^2) \tag{29}$$

and allocate it to ys_t. Otherwise, we allocate q_t to ys_t, draw d_t according to

$$d_t \sim N(d_t|\mu_{dt}, \sigma_1^2) \tag{30}$$

and allocate it to yd_t. We have thus two sets of endogenous variables mixing observed and simulated values. Using the noninformative prior (28), we find that the conditional posterior density of the parameters is

$$\begin{aligned} \varphi(\gamma_1|\sigma_1^2, yd) &= f_N(\gamma_1|(Z_1'Z_1)^{-1}Z_1'yd, \sigma_1^2) \\ \varphi(\sigma_1^2|yd) &= f_{I\gamma 2}(\sigma_1^2|, yd'yd - yd'Z_1(Z_1'Z_1)^{-1}Z_1'yd, T-1) \end{aligned} \tag{31}$$

for the demand regime, and similarly for the supply regime:

$$\begin{aligned} \varphi(\gamma_2|\sigma_2^2, ys) &= f_N(\gamma_2|(Z_2'Z_2)^{-1}Z_2'ys, \sigma_2^2) \\ \varphi(\sigma_2^2|ys) &= f_{I\gamma 2}(\sigma_2^2|, ys'ys - ys'Z_2(Z_2'Z_2)^{-1}Z_2'ys, T-1). \end{aligned} \tag{32}$$

Starting values for the Gibbs sampler can be easily chosen. We can suppose that the market is in equilibrium and thus impose $yd = ys = q$ where q is the $(T-1)$-dimensional vector of observed quantities q_t. We then compute the OLS estimators $\hat{\gamma}_i = (Z_i'Z_i)^{-1}Z_i'q$ and $\hat{\sigma}_i^2 = q'q - q'Z_i(Z_i'Z_i)^{-1}Z_i'q$, $i = 1, 2$, and use these estimates as starting values. Even if these estimates are not the best possible starting values, they have the advantage of being feasible as they assume the model is near equilibrium.

3.3 Identification issues

The Gibbs algorithm hides a potential identification problem which is quite apparent when we consider the direct likelihood function. Let us write it using formulation (22) of the model. As δ_t can be only either 0 or 1, its role is just to select observations. Consequently, we have the following equivalence:

$$\begin{aligned} L(q;\theta) &= \prod_t [\delta_t f_N(q_t|\mu_{dt},\sigma_1^2) + (1-\delta_t) f_N(q_t|\mu_{st},\sigma_2^2)] \\ &= \prod_{t|\delta_t=1} f_N(q_t|\mu_{dt},\sigma_1^2) \prod_{t|\delta_t=0} f_N(q_t|\mu_{st},\sigma_2^2). \end{aligned} \tag{33}$$

For identification, the information matrix associated to this likelihood function must be invertible. We write the identification conditions in the following theorem.

Theorem 1 *A necessary and sufficient identification condition for model (22) is that there are at least as many observations as parameters in each regime, provided the regressors are not collinear.*

Proof: For a given vector $\delta = [\delta_2, \ldots, \delta_T]$, (33) is the likelihood function of a Gaussian regression model with unequal variances. The information matrix for the regression parameters is given by

$$I(\gamma_1,\gamma_2)) = \begin{pmatrix} \sigma_1^{-2} \sum_{t|\delta_t=1} z_{1t}z_{1t}' & 0 \\ 0 & \sigma_2^{-2} \sum_{t|\delta_t=0} z_{2t}z_{2t}' \end{pmatrix}.$$

In the absence of collinearity, this matrix is regular whenever $\sum \delta_t \geq k_1$ or $\sum \delta_t \leq T - k_2$, where k_i the number of regressors in each regime.

□

The integrability of the posterior density requires the model to be identified over the whole feasible domain of integration. Obviously we need an informative prior to reach integrability. We propose to modify (28) into

$$\varphi(\theta) \propto \sigma_1^{-2}\sigma_2^{-2}\mathbb{I}(k_1 \leq \sum \delta_t \leq T - k_2) \tag{34}$$

This prior modifies the Gibbs sampler by just introducing a supplementary rejection step.

3.4 Computational efficiency of the Gibbs sampler

Lubrano (1985) has analyzed the static GTZ model using importance sampling to integrate directly the posterior density. Do we gain computational efficiency by using a Gibbs sampler with data augmentation? We performed a small experiment on simulated data to compare our algorithm with a a Metropolis algorithm applied directly to the likelihood function (21). The data generating process is

$$\begin{aligned} d_t &= \rho_1 q_{t-1} + \beta_{10} + x_{1t}\beta_{11} \\ s_t &= \rho_2 q_{t-1} + \beta_{20} + x_{2t}\beta_{21} \\ q_t &= \text{Min}(s_t, d_t) + \epsilon_t. \end{aligned} \tag{35}$$

The variable x_{1t} is generated by the autoregressive process $x_{1t} = 0.85x_{1t-1} + u_t$ where $u_t \sim N(0,1)$ and $x_0 = 3.5$, and x_{2t} is generated by the same process, but independently of x_{1t}. The error term ϵ_t is generated as a $\mathrm{N}(0, \sigma^2)$ with $\sigma^2 = 0.05$. We generated 250 observations using the parameter values indicated in Table 1; 128 are in the first regime and 122 in the second regime. The value of the parameters imply a R^2 of 0.95 in each regime.

Table 1: Gibbs sampler with data augmentation
versus direct integration using Metropolis

	ρ_1	β_{10}	β_{11}	ρ_2	β_{20}	β_{21}	σ_1^2	σ_2^2
True values	0.60	-2.00	1.00	0.40	7.00	-1.50	0.050	0.050
Starting values	0.81	-0.67	0.35	0.82	1.75	-0.38	0.18	0.17
Gibbs	0.63	-2.08	1.02	0.45	6.46	-1.39	0.045	0.052
	0.025	0.15	0.059	0.032	0.33	0.071	0.0061	0.0065
Metropolis	0.63	-2.02	1.00	0.44	6.58	-1.42	0.045	0.053
	0.0099	0.065	0.025	0.013	0.15	0.033	0.0028	0.00030

In the Gibbs and Metropolis blocks, the first row contains the posterior means and the second the posterior standard deviations.

Let us first analyse this simulated sample using the a Gibbs sampler with data augmentation. As starting values, we used the OLS estimator assuming equilibrium. Table 1 indicates that this procedure gives starting values which are rather far from the true generating values, with however a correct sign. They induce a feasible sample separation. We used 10 000 draws plus 1 000 for warming up the chain. Convergence was checked using CUMSUM graphs. The Gibbs sampler took 1.92 seconds to get the results presented in the first part of Table 1. The total variance of this Monte Carlo experiment,[2] computed using a parametric estimate of the spectral density at zero to take into account the correlation among the draws, is 4.898E-8. If we raise the number of draws to 40 000, the total Monte Carlo variance goes down to 1.262E-8 for a total computing time of 7.03 seconds.

It was much more difficult to achieve a comparable result with the direct approach using a Metropolis algorithm to integrate the posterior density. The candidate function is a Student with 4 degrees of freedom. A first round of 10 000 draws and a random walk algorithm, with the same starting values as for the Gibbs sampler plus a diagonal covariance matrix determined so as to get t-statistics equal to one, took 2.09 seconds with a Metropolis rejection rate of 99.5%. A second round of the random walk algorithm, using the results of the previous round for re-calibrating the candidate density, took 2.00 seconds with a rejection rate of 46%. The total variance of this experiment is 6.111E-7. We needed a third iteration, with this time an independent Metropolis (2.14 seconds and a rejection rate of 71%) to get the smaller Monte Carlo variance of 2.78E-8.

[2]The variance of a MCMC experiment is defined by formula (3.50) of Bauwens, Lubrano, and Richard (1999) for one parameter. The values reported in the text are obtained by adding the variances computed for each parameter.

However, the graphs of the marginal densities were not very smooth, which is a clear sign of a lack of convergence. A fourth iteration with the independent Metropolis had a rejection rate of 42% and a Monte Carlo total variance of 7.074E-9.

On this particular simulated sample, the Metropolis algorithm applied to the likelihood function needed four iterations to achieve convergence. The Gibbs sampler converged with only 10 000 draws and 1.92 seconds, compared to the 40 000 of the Metropolis procedure and a total computer time of 8.37 seconds. However, once convergence is reached, the direct approach has of course a smaller Monte Carlo variance, 7.074E-9 against 1.262E-8 for the Gibbs sampler with 40 000 draws, because the integration problem is smaller in the direct approach. Table 1 also indicates that posterior standard deviations found with the direct approach and the Metropolis algorithm are on average half of those found by data augmentation and the Gibbs sampler. We computed classical OLS estimates using the optimal sample separation found by the Metropolis algorithm as an approximation for the MLE. The corresponding classical standard deviations were on average 6% lower than the Bayesian standard deviations found by the Gibbs sampler and consequently much larger than the posterior standard deviations found by the Metropolis algorithm. We can conclude that the Metropolis algorithm has underestimated the posterior standard deviations because of its large rejection rate.

4 Model comparison

In any empirical application, one must be able first to check a model against directions of misspecification and second to compare final alternative specifications. Both procedures are closely related to predictive analysis in that they involve the computation of the marginal likelihood or transformations of the marginal likelihood. This problem has received much attention in the Bayesian literature, starting with the posterior odds ratio of Jeffreys (1961). However, posterior odds became increasingly criticized as they are not well defined when using improper or diffuse priors. The null hypothesis is never rejected when one is diffuse on the parameters of the alternative hypothesis. The current approach (see e.g. Newton and Raftery (1994), Chib (1995), Gelman and Meng (1996), Spiegelhalter, Best, Carlin, and van der Linde (2002)) tends to define criteria based on the MCMC output. We adopt this approach for our specification search on the GTZ model in the next section.

The BIC (Bayesian information criterion) is a large sample approximation of the Bayes factor. It penalizes a measure of fit by a measure of complexity, the latter being defined simply by the number of parameters. This measure of complexity is not very relevant for models with latent variables where the number of parameters is hard to define. This is also the case in non-linear models because a non-linear model is intuitively more complex than a linear model with the same number of parameters. Spiegelhalter, Best, Carlin, and van der Linde (2002) developed the deviance information criterion (DIC) which penalizes a measure of fit by a measure of complexity. This criterion is based on the classical deviance defined as

$$D(\theta) = -2\log f(y|\theta) + 2\log f(y), \tag{36}$$

where $f(y|\theta)$ is the likelihood function and the last term plays the role of a normalising constant.

It is a function of the data alone and therefore is irrelevant for model comparison. For an absolute measure of fit, $f(y)$ represents the saturated model and $D(\theta)$ is thus the saturated deviance.

A Bayesian measure of fit is given by the posterior expectation of the deviance:

$$\overline{D} = \mathrm{E}_{\theta|y} D(\theta). \tag{37}$$

As the deviance is minus twice the log-likelihood function, a smaller posterior deviance implies a better fit.

Spiegelhalter *et al* (2002) use a decision theory argument to define a measure of complexity p^* as the difference between the posterior expectation of the deviance and the deviance computed at the posterior expectation of the parameters (denoted by $\bar{\theta}$), i.e.

$$p^* = \overline{D} - D(\bar{\theta}). \tag{38}$$

This equation can be rearranged so as to redefine $\overline{D} = D(\bar{\theta}) + p^*$, which shows that the posterior deviance can be seen as a classical plug-in measure of fit plus a penalty term. This qualifies it more as a measure of adequacy than a pure measure of fit. One can also relate this measure of complexity to the total number of parameters p of the model. If we approximate the posterior density by a normal density centered on the MLE estimator $\hat{\theta}$ and suppose that the prior is dominated by the sample, then $D(\theta) \simeq D(\hat{\theta}) + \chi^2(p)$ (formula (18) in Spiegelhalter *et al.* 2002). One can then show that $\mathrm{E}_{\theta|y}[D(\theta) - D(\hat{\theta})] \simeq p$.

The deviance information criterion (DIC) is defined as the sum of the posterior deviance plus the above measure of complexity:

$$DIC = \overline{D} + p^*. \tag{39}$$

However, using the fact that $\overline{D} = D(\bar{\theta}) + p^*$, we can give another expression for the DIC which immediately relates it to the classical AIC criterion:

$$DIC = D(\bar{\theta}) + 2p^*. \tag{40}$$

We summarize in Table 2 three criteria to compare models, and their asymptotic approximations. They are the Bayes factor, which receives the BIC as an approximation, the deviance information

Table 2: Bayesian information criterion
and large sample approximations

Bayesian criteria	Large sample approximations
Bayes Factor	$BIC = D(\hat{\theta}) + p \log T$
DIC	$AIC = D(\hat{\theta}) + 2p$
$\overline{D}$	$D(\hat{\theta}) + p$

$\hat{\theta}$ represents the maximum likelihood estimate of θ.

criterion DIC that can be approximated by the AIC, and finally the posterior deviance $\overline{D}$, which is the less penalized criterion. We use this last criterion for selecting models in the empirical application.

These three criteria can be easily computed from the draws of a MCMC algorithm.[3] In the case of the GTZ model, we have a direct expression of the likelihood function, so the computation of the DIC is simple. In a pure latent variable model, the situation is more complex because the latent variables have to be replaced by their simulations, see for instance Berg, Meyer, and Yu (2004) for stochastic volatility models.

5 Credit rationing in Poland

Stiglitz and Weiss (1981) were the first to show that credit rationing can occur even at equilibrium because the credit market is a market where the expected quality of the product is a function of its price. At a given interest rate, some borrowers will receive a loan and others not. Those not receiving a loan would be ready to pay more, but they are thought of being too risky. On the other side, banks are not tempted by lowering rates when they have an excess of liquidity because they may attract risky borrowers from other banks. Consequently, the short side of the market will prevail.

Since the pioneering work of Laffont and Garcia (1977), many disequilibrium models have been estimated for credit markets. They however mainly concern emerging or countries in transition. Some of the most recent references are Kim (1999) for Korea, Barajas and Steiner (2001) for Latin America, Ikhide (2003) for Namibia, Shen (2002) for Taiwan, and Hurlin and Kierzenkowski (2002, 2003) for Poland. We have chosen this last case for illustrating our estimation procedures.

The Polish credit market is interesting because it represents something like a textbook case as underlined by Hurlin and Kierzenkowski (2003). Poland experienced large foreign liquidity inflows between 1995 and 1998 which induced a boom in credit supply. The situation changed suddenly in 1999 because of the instability of the real activity. Banks profitability fell down, they contracted their credit supply and tried to invest more in riskless assets. This suggests an excess credit supply before 1999 and an excess demand after that date.

Previous studies of the Polish credit market are based on monthly data covering the period 1994:02-2002:02. There has been a major change in data definition and construction in 2002 so as to harmonize the definition of money and its counterparts with the standards binding upon the member states of the European System of Central Banks. Thus, as of end-March 2002 the consolidated balance sheets of the banking system are presented in a new format. An effort has been made by the National Bank of Poland to retropolate the data so as to obtain homogenous time series in the new format from 1996:12. The National Bank of Poland gave us a monthly data set of monetary series covering the period 1996:12-2003:11, representing 84 observations, completed by monthly macroeconomic series covering a longer period.

[3] One way of approximating the Bayes factor is to use the harmonic mean of the posterior draws from the MCMC output.

5.1 Data description

Our dependent variable represents the loans up to one year extended to the Polish corporate sector. We have observations both on zloty loans and on loans in foreign currencies. The latter represent on average 18.50% of total loans on the period, with large fluctuations. Figure 1 displays the series of zlotys loans and of total loans (zloty+foreign currency), denominated in millions of zlotys. We

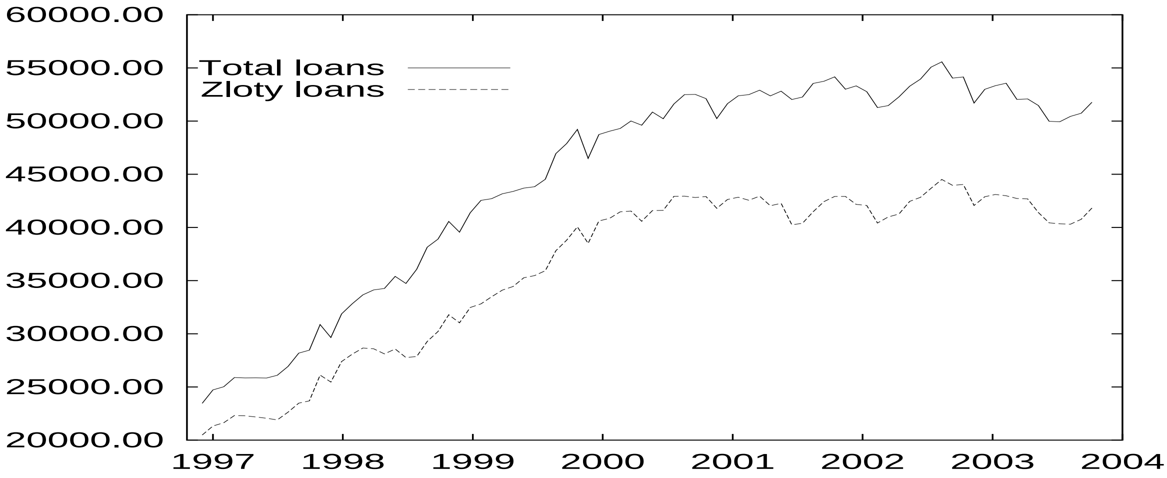

Figure 1: Short term loans to the corporate sector in Poland (in million of zlotys)

can notice the important increases of the loan series until 2000, followed by a stabilization which is more pronounced for zlotys loans than for total loans.

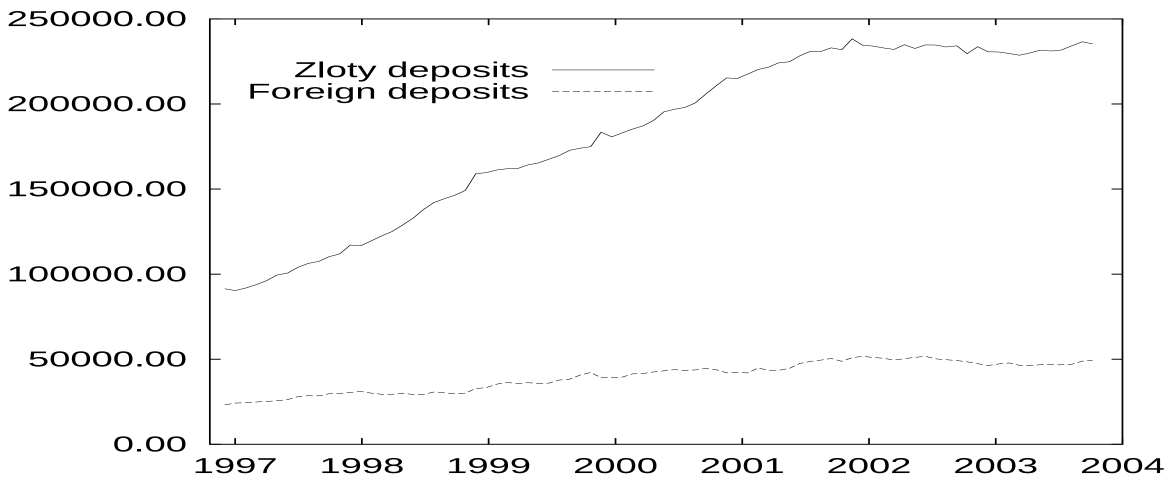

Figure 2: Zloty and foreign deposits in Polish banks (in million of zlotys)

Figure 2 displays the evolution of foreign and domestic deposits in Polish banks. The inflow of foreign deposits increased until 2002 and remained stable afterwards. The increase in zloty deposits is much more pronounced before 2002 and stopped afterwards.

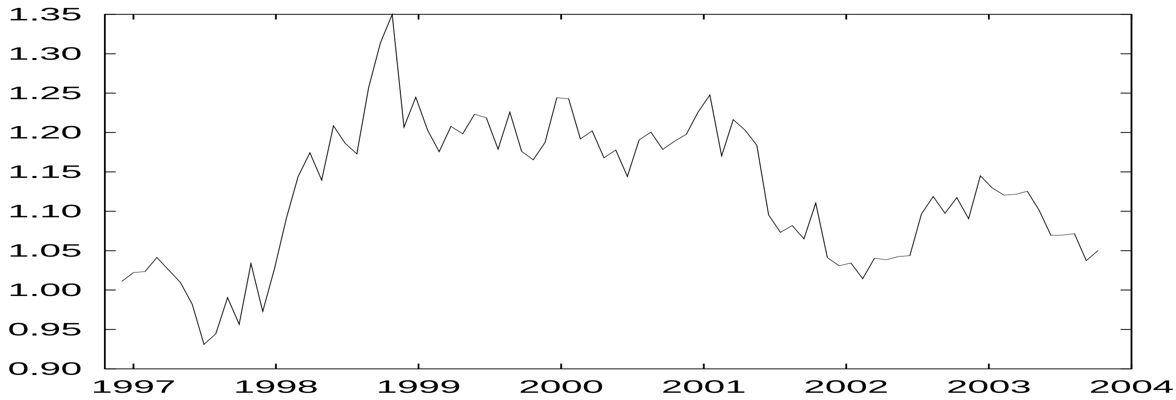

Figure 3: Short term lending ratio for foreign deposits

The evolution of the profitability of banks is well depicted in Figure 3 which displays the short term lending ratio[4] using foreign deposits: a large increase till 1999 and a stabilization followed by a decrease. This graph is well in accordance with the above description of the Polish credit market. However, if we take the total deposits to compute the short term lending ratio, we find a high and stable ratio till 2000 followed by a large and regular decrease after that date.

5.2 Model specification

The supply equation in usual credit models is determined mainly by resources and interest rate variables. There are essentially two interest rates: the lending rate (LR) and the intervention rate of the central bank (IR). Loan supply should depend positively on the lending rate and on the deposits (DEP) and negatively on the intervention rate. However, the excess of liquidity present at least at the beginning of the period may have reduced greatly the effect of interest rates on credit supply. We model short term credit supply. We suppose that banks had to make two types of arbitrages between different allocations. Firstly, instead of supplying short term credit, they may prefer to supply longer term credits. Secondly, if they judge that the market is too risky, they may prefer to invest in riskless assets like treasury bills. The variable OLT is equal to the ratio of total loans longer than one year to total extended loans. The variable OTB represents the ratio between treasury bills bought by the banking system and the total banking assets. Both variables should have a negative impact on credit supply. Consequently, the supply plan μ_{st} of banks is specified by

$$\mu_{st} = a_1 + a_2 LQ_{t-1} + a_3 OLT_t + a_4 DEP_t + a_5 LR_t + a_6 IR_t + a_7 OTB_t \qquad (41)$$

[4]The lending ratio is defined as the ratio between loans and deposits. It is a traditional quantitative measure of profitability for banks because is relates extended loans to the bank resources.

where all variables are taken in logarithms. For the interest rate variables, we took $\log(100 + R)$. As one can suppose that banks have a short term horizon in term of lending plans, we have chosen a lag of 1 for LQ, the amount of loans (the dependent variable).

Loan demand by firms is usually said to depend more on profit anticipations than on interest rates. The lagged industrial production IP is usually taken as a proxy for future activity. Finally firms operate an arbitrage between short term and long term loans which may depend on the structure of interest rates. Consequently we specify the demand plan μ_{dt} made by firms by

$$\mu_{dt} = b_1 + b_2 LQ_{t-12} + b_3 OLT_t + b_4 IP_{t-12}. \tag{42}$$

As we can suppose that firms make plan at a longer horizon than banks, we have chosen a lag of 12 for both LQ and IP.

Market clearing is specified by the GTZ rule

$$LQ_t = \text{Min}(\mu_{st}, \mu_{dt}) + \epsilon_t, \tag{43}$$

where ϵ_t may have a different variance in each regime. We have decided to model the credit market for total short term loans. Thus, LQ represents loans in zlotys and loans in foreign currencies up to one year extended to the Polish corporate sector. We have made specific hypotheses concerning the lag structure of the two equations. This is of course a debatable question that needs a clear evaluation.

5.3 Specification search and inference results

We made a specification search, trying various formulations for the supply equation. The demand equation (42) seems to be rather robust and was kept the same during the search. We started with a supply equation based on (41), with the amount of foreign deposits ($DEPF_t$) instead of the total deposits (DEP_t). We obtained the following results together with a posterior deviance equal to 211.3:

$$\begin{aligned} LS_t &= \underset{[2.27]}{7.07} + \underset{[0.17]}{0.33} LQ_{t-1} - \underset{[0.25]}{0.64} OLT_t + \underset{[0.076]}{0.018} DEPF_t \\ &\quad + \underset{[0.64]}{0.32} LR_t - \underset{[0.62]}{0.37} IR_t + \underset{[0.0094]}{0.012} OTB_t \\ LD_t &= \underset{[0.40]}{0.95} + \underset{[0.023]}{0.67} LQ_{t-12} - \underset{[0.28]}{1.57} OLT_t + \underset{[0.055]}{0.17} IP_{t-12}. \end{aligned}$$

Posterior standard deviations are indicated between brackets below the posterior means. The coefficients of the demand equation have all the correct sign and the posterior $p-$value of zero is always lower than 1%. The coefficients of the supply equation have also the correct sign, but most of them have a very large posterior standard deviation. Despite a large number of draws[5], CUMSUM graphs indicated that convergence for this equation is not well established. We eliminated successively the interest rate variables to get a model with a posterior deviance of 209.4. In this

[5] 100 000 draws for this first specification, and only 20 000 for the final model.

reduced specification, the treasury bill variable (OTB) has the wrong sign, but for this coefficient zero has a posterior $p-$value of 20%. After eliminating this variable, we get our final specification:

$$\begin{aligned} LS_t &= \underset{[1.00]}{3.43} + \underset{[0.11]}{0.59} LQ_{t-1} - \underset{[0.14]}{0.26} OLT_t + \underset{[0.054]}{0.086} DEPF_t \\ LD_t &= \underset{[0.41]}{0.64} + \underset{[0.026]}{0.67} LQ_{t-12} - \underset{[0.33]}{1.60} OLT_t + \underset{[0.061]}{0.20} IP_{t-12} \end{aligned}$$

with a posterior deviance of 208.1 to be compared to 211.3 for the starting specification. The posterior expectations of the variance of the error term in the two regimes are almost equal: $\mathrm{E}(\sigma_s^2|y) = 0.00022$ and $\mathrm{E}(\sigma_d^2|y) = 0.00023$.

The short term loan supply depends positively on foreign deposits, but with a long run elasticity of 0.21. The decision of supplying long term loans as an alternative has a stronger and negative influence on short term loan supply, with an elasticity of -0.63. The mean lag of reaction is one month and a half.

The short term loan demand is mainly determined by an arbitrage decision between short term and long term loans with a strong elasticity of -4.85, but a mean lag of two years. The other decision variable is the prediction made on economic activity with an elasticity of 0.61 and a mean lag of three years.

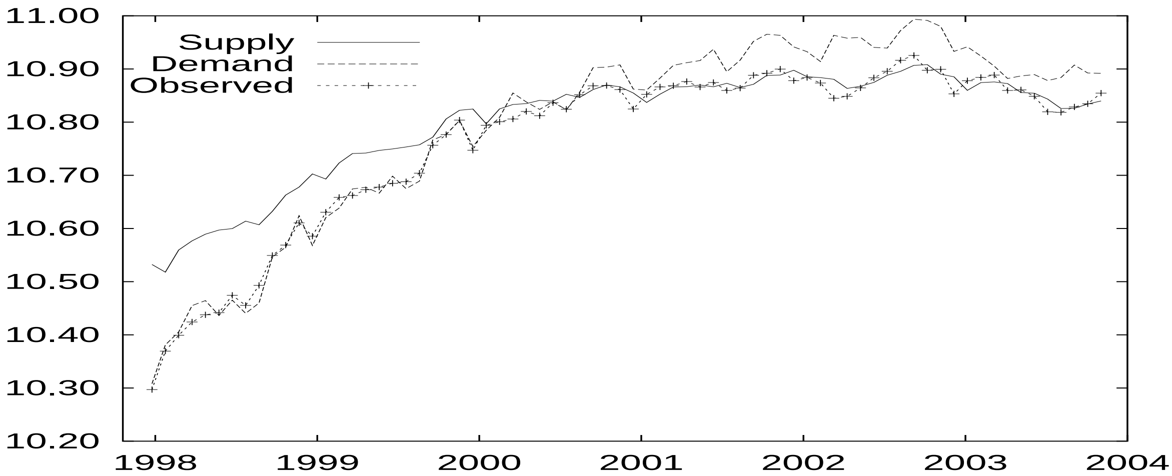

Figure 4: Estimated loan supply and loan demand with the final GTZ model

At the posterior expectation of the final model, there are 43 observations in the demand regime and 29 in the supply regime, see Figure 5. The obtained regime separation is in accordance with the stylized facts described in the introduction of this section. We have a large excess supply till the mid of 1999 (see Figure 4), corresponding to a slightly later date than that indicated in the introduction. There is a period of large excess demand after the beginning of 2001. Between these two dates, we have a transition period with an alternation of excess demand and excess supply.

Our final model gives a clear picture of the Polish credit market. Short term credit supply and demand are both governed by an arbitrage decision between short term and longer term cred-

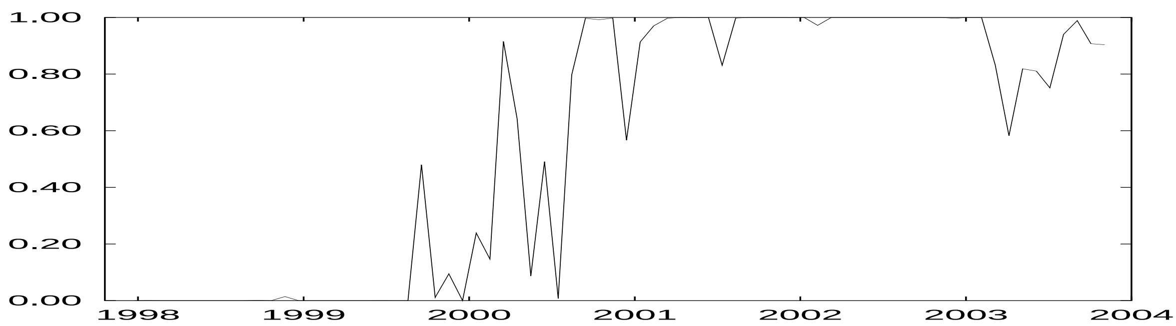

Figure 5: Probabilities of excess demand with the final GTZ model

its. Supply additionally depends on deposit resources and demand on anticipated activity. The main difference between demand and supply is in term of speed of reaction to variation of these variables: a rather quick reaction for suppliers and a much slower reaction of demanders.

These final results were obtained with 20 000 draws of the Gibbs sampler (5 000 draws were discarded for warming up the chain). Starting values are the OLS estimates on the entire sample for both equations (ignoring thus sample separation). CUMSUM graphs revealed no problem of convergence. We used the partially diffuse prior (34). However the identification restriction implied by this prior was never activated by the Gibbs sampler.

5.4 Alternative specifications

There is a big difference in the dynamics of the two equations. This is mainly due to the fact that we impose a lag of one in the supply equation and a lag of twelve in the demand equation. Otherwise, the two autoregressive coefficients are comparable (0.59 and 0.67 in the final model). The regime separation appears to be very sensitive to this choice. Choosing a lag of one in both equations leads to a counterintuitive sample separation. This choice can be easily discarded because it gives a posterior deviance of 239.5, much higher than the value 208.1 of the final model. We note that Hurlin and Kierzenkowski (2003) also experienced difficulties in getting a correct sample separation with their totally static model.

The specification search we presented was simple because $p-$values indicated clearly which variables to delete. This simplicity comes from that fact that the sample separation was fairly stable during the search. When we start the specification search including also zloty deposits in the supply equation, the obtained sample separation is very different and there is no immediate direction for simplification despite the fact that the posterior deviance of 211.2 is greater than that of our final model.

Finally, let us report some variants of our final model, by replacing the foreign deposit variable in the supply equation either by zloty deposits or by total deposits. This gives equivalent results in term of posterior deviance. Using zloty deposits ($DEPZ_t$) instead of foreign deposits ($DEPF_t$)

in the supply equation, we get

$$
\begin{aligned}
LS_t &= \underset{[0.91]}{3.47} + \underset{[0.12]}{0.45}\, LQ_{t-1} - \underset{[0.16]}{0.48} OLT_t + \underset{[0.078]}{0.18}\, DEPZ_t \\
LD_t &= \underset{[0.42]}{0.60} + \underset{[0.029]}{0.68}\, LQ_{t-12} - \underset{[0.32]}{1.55} OLT_t + \underset{[0.062]}{0.20}\, IP_{t-12}
\end{aligned}
$$

with a posterior deviance of 208.4. If we use total deposits (DEP_t) in the supply equation, we have

$$
\begin{aligned}
LS_t &= \underset{[0.75]}{3.28} + \underset{[0.11]}{0.46}\, LQ_{t-1} - \underset{[0.15]}{0.46} OLT_t + \underset{[0.073]}{0.19}\, DEP_t \\
LD_t &= \underset{[0.42]}{0.56} + \underset{[0.030]}{0.68}\, LQ_{t-12} - \underset{[0.33]}{1.50} OLT_t + \underset{[0.063]}{0.21}\, IP_{t-12}
\end{aligned}
$$

with a posterior deviance of 208.0. It appears thus that the results are rather robust with respect to the choice of the deposit variable in the supply equation. This last specification can be in fact the preferred one since it has the smallest posterior deviance. It is also the most coherent because the supply equation relates total short term credits to total deposits.

6 Conclusion

The data augmentation principle is a nice and simple solution for Bayesian inference in latent variable models as long as there is no dynamics. We have reviewed in Section 2 the treatment of the static Tobit model and of the static disequilibrium model of Maddala and Nelson (1974). The dynamic case causes special problems due to the fact that it is not correct to simulate the latent variables conditionally on the observations. Special algorithms have to be implemented, such as the scan sampler of de Jong (1997). Instead of considering a dynamic version of the model of Maddala and Nelson (1974), we propose to use the alternative specification suggested by Ginsburgh, Tishler, and Zang (1980). We have shown that the implementation of the data augmentation principle for this model is both simple and efficient. We use the the posterior deviance criterion of Spiegelhalter, Best, Carlin, and van der Linde (2002) to compare models.

In our application to the Polish credit market, we recover some stylized facts of the Polish economy. We also underline that a correct dynamic starting specification is essential for conducting a valuable specification search. An incorrect dynamic model or the inclusion of too many variables may lead to a model which generates a counterintuitive sample separation. This issue is at the heart of disequilibrium models which are in this respect very different from the usual linear equilibrium models.

References

BARAJAS, A., AND R. STEINER (2001): "Credit stagnation in Latin America," IMF Working Papers, No 02/53.

BAUWENS, L., M. LUBRANO, AND J.-F. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.

BERG, A., R. MEYER, AND J. YU (2004): “Deviance information criterion for comparing stochastic volatility models,” *Journal of Business and Economic Statistics*, 22(1), 107–120.

CHIB, S. (1992): “Bayes inference in the tobit censored regression model,” *Journal of Econometrics*, 51(1-2), 79–99.

——— (1995): “Marginal Likelihood from the Gibbs Output,” *Journal of the American Statistical Association*, 90, 1313–1321.

DAGENAIS, M. (1982): “The Tobit model with serial correlation,” *Economic Letters*, 10, 263–267.

DE JONG, P. (1997): “The scan sampler,” *Biometrika*, 84, 929–937.

DRÈZE, J. H., C. BEAN, J.-P. LAMBERT, F. METHA, AND H. R. SNEESSENS (eds.) (1991): *Europe’s Unemployment Problem*. MIT Press, Cambridge, Mass.

GELMAN, A., AND X. MENG (1996): “Model checking and model improvement,” in *Practical Markov Chain Monte Carlo*, ed. by W. Gilks, S. Richardson, and S. Spiegelhalter, pp. 189–201. Chapman and Hall, London.

GINSBURGH, V., A. TISHLER, AND I. ZANG (1980): “Alternative estimation methods for two regime models,” *European Economic Review*, 13, 207–228.

HURLIN, C., AND R. KIERZENKOWSKI (2002): “A theoretical and empirical assessment of the bank lending channel and loan market disequilibrium in Poland,” National Bank of Poland, working paper No 22.

——— (2003): “Credit market disequilibrium in Poland: Can we find what we expect? Nonstationarity and the Min condition,” University of Michigan, William Davidson Institute Working Papers No 581.

IKHIDE, S. (2003): “Was there a credit crunch in Namibia between 1996-2000?,” *Journal of Applied Economics*, 6(2), 269–290.

JACQUIER, E., N. POLSON, AND P. ROSSI (1994): “Bayesian analysis of stochastic volatility,” *Journal of Business and Economic Statistics*, 12(4), 371–417.

JEFFREYS, H. (1961): *The Theory of Probability*. Clarendon Press, Oxford, 3rd edn.

KIM, H. (1999): “Was credit channel a key monetary transmission mechanism following the recent financial crisis in the Republic of Korea?,” The World Bank, Policy Research Working Paper 3003.

KIM, S., N. SHEPHARD, AND S. CHIB (1998): “Stochastic volatility: likelihood inference and comparison with ARCH models,” *Review of Economic Studies*, 65(3), 361–393.

LAFFONT, J.-J., AND R. GARCIA (1977): “Disequilibrium econometrics for business loans,” *Econometrica*, 45(5), 1187–1204.

LAROQUE, G., AND B. SALANIE (1993): "Simulation based estimation of models with lagged latent variables," *Journal of Applied Econometrics*, 8(0) Suppl., S119–S133.

LEE, L. F. (1997): "A smooth likelihood simulator for dynamic disequilibrium models," *Journal of Econometrics*, 78(2), 257–294.

——— (1999): "Estimation of dynamic and ARCH tobit models," *Journal of Econometrics*, 92(2), 355–390.

LUBRANO, M. (1985): "Bayesian inference in switching regression models," *Journal of Econometrics*, 29, 69–85.

MADDALA, G. (1987): *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, New-York.

MADDALA, G., AND F. NELSON (1974): "Maximum likelihood methods for models of markets in disequilibrium," *Econometrica*, 42(6), 1013–1030.

MANRIQUE, A., AND N. SHEPHARD (1998): "Simulation-based likelihood inference for limited dependent processes," *Economerics Journal*, 1, C174–C202.

NEWTON, M. A., AND A. E. RAFTERY (1994): "Approximate Bayesian inference with the weighted likelihood bootstrap," *Journal of the Royal Statistical Society*, Ser. B, 56, 3–48.

RICHARD, J.-F. (1980): "C-type distributions and disequilibrium models," Paper presented at the Conference "Disequilibrium models" held in Toulouse.

ROBERT, C., AND G. CASELLA (1999): *Monte Carlo Statistical Methods*, Springer Texts in Statistics. Springer Verlag, New York.

SHEN, C.-H. (2002): "Credit rationing for bad companies in bad years: Evidence from bank loan transaction data," *International Journal of Finance and Economics*, 7(3), 261–78.

SNEESSENS, H. (1985): "Two alternative stochastic specifications and estimation methods for quantity rationing models: a Monte Carlo comparison," *European Economic Review*, 29, 111–136.

SPIEGELHALTER, D. J., N. G. BEST, B. P. CARLIN, AND A. VAN DER LINDE (2002): "Bayesian measures of complexity and fit," *Journal of the Royal Statistical Society*, Ser. B, 64, 583–639.

STIGLITZ, J., AND A. WEISS (1981): "Credit rationing in markets with imperfect information," *The American Economic Review*, 71(3), 393–410.

TANNER, M. A., AND W. H. WONG (1987): "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, 82(398), 528–540.

TOBIN, J. (1958): "Estimation of relationships for limited dependent variables," *Econometrica*, 26(1), 24–36.

WEI, S. X. (1999): "A Bayesian approach to dynamic tobit models," *Econometric Reviews*, 18(4), 417–439.

ZANGARI, P., AND H. TSURUMI (1996): "A Bayesian analysis of censored autocorrelated data on exports of Japanese passenger cars to the United States," in *Bayesian computational methods and applications*, ed. by C. Hill, vol. 11. (Part A) of *Advances in Econometrics*, pp. 111–143. JAI Press, Greenwich, Conn. and London.

Département des Sciences Économiques
de l'Université catholique de Louvain
Institut de Recherches Économiques et Sociales

Place Montesquieu, 3
1348 Louvain-la-Neuve, Belgique

ISSN 1379-244X D/2006/3082/027