Estimating Default and Recovery Rate Correlations

JIŘÍ WITZANY[1]

Abstract: The paper analyzes a two-factor credit risk model allowing to capture default and recovery rate variation, their mutual correlation, and dependence on various explanatory variables. At the same time, it allows computing analytically the unexpected credit loss. We propose and empirically implement estimation of the model based on aggregate and exposure level Moody's default and recovery data. The results confirm existence of significantly positive default and recovery rate correlation. We empirically compare the unexpected loss estimates based on the reduced two-factor model with Monte Carlo simulation results, and with the current regulatory formula outputs. The results show a very good performance of the proposed analytical formula which could feasibly replace the current regulatory formula.

AMS/JEL classification: G20, G28, C51

Keywords: credit risk, Basel II regulation, default rates, recovery rates, correlation

1 Introduction

The goal of this paper is to study a viable alternative of the Basel II regulatory formula (see Basel, 2006). Such a formula should provide a sufficiently robust estimate of unexpected losses on a portfolio of banking credit exposures to be covered by capital. The capital requirement of a bank in the current Basel II Internal Rating Based (IRB) approach is calculated as the unexpected loss (*UL*) less the expected loss (*EL*)

$$C = UL - EL = (UDR - PD) \times LGD \times EAD \tag{1}$$

decomposed into the product of the unexpected default rate (*UDR*) less the expected default rate (*PD*), loss given default (*LGD*), and the exposure at default (*EAD*). The calculation is done on the level of each individual exposure, but the total should correspond to portfolio unexpected credit loss on the 99.9% probability level. The unexpected default rate (*UDR*) that is calculated as a regulatory function of *PD*, asset correlation ρ (set by the regulation

[1]University of Economics in Prague, Winston Churchill Sq. 4, Prague 3, Czech Republic, E-mail: jiri.witzany@vse.cz
This research has been supported by the Czech Science Foundation Grant P402/12/G097 "Dynamical Models in Economics"

depending on the asset class and *PD*). The formula is based on the assumption that the event of default is driven by a normally distributed variable. Moreover, the account level risk driving factor is decomposed into a single normally distributed systematic factor and into an independent normally distributed idiosyncratic factor (Vasicek, 1987). While *UDR* is calculated by a relatively sophisticated model, the regulatory approach simplifies the analysis of the remaining two parameters just vaguely requiring that the estimates reflect downturn economic conditions and possible correlations with the rate of default (BCBS, 2005). This approach is not changed by the latest Basel III regulatory reform (BCBS, 2010).

Witzany (2010a) provides an overview of several single factor models (Frye, 2000a, Pykhtin 2003, and Tasche, 2004,or Gordy, 2003) and analyzes certain surprising effects of the regulatory formula that are caused by the fact that any single factor model cannot properly capture correlation between the default and recovery rates. In fact, it has been empirically shown in a number of papers by Altman et al. (2004, 2007), Gupton et al. (2000), Frye (2000b, 2003), Acharya et al. (2007), or Seidler (2009) that there is not only a significant systematic variation of recovery rates but, moreover, a negative correlation between the default and recovery rates, or equivalently a positive correlation between the rates of default and the rates of loss given default.

The correlation can also be analyzed through dependence of the default and recovery rates on common macroeconomic factors. For example, Jacobson et al. (2005) or De Graeve et al. (2008) examine the relationship between the default rate and macroeconomic factors like output, inflation, and interest rates. On the other hand, Casselli et al. (2008), Bellotti and Crook (2009), or Belyaev et al. (2012) confirm a significant relationship between LGD and a number of macroeconomic variables like GDP growth, unemployment rate, household consumption, or directly the default rate itself.

Consequently, if the LGD parameter does not reflect the systematic correlation with the default rate then the regulatory formula might significantly underestimate the potential unexpected losses.

Witzany (2011) estimates default, recovery and mutual default-recovery rate correlations based on a two-systematic-factor model. In this study, the recovery rate can have any parametric or nonparametric distribution. This generality makes the estimation procedure more difficult and the subsequent unexpected losses can be estimated using a Monte Carlo simulation only. In the presented study we will restrict ourselves to the two-systematic-factor model of Rosch and Scheule (2009) where the inverse Probit transformed recovery rates are

assumed to be normally distributed and the event of default is driven, as usual, by a normally distributed variable. The great advantage of this restriction is that it leads to a relatively simple analytical formula estimating consistently downturn LGD where the inputs are: expected LGD, recovery rate systematic factor loading, and a default – recovery rate correlation coefficient. Consequently, the formula is a viable candidate that could serve as a regulatory downturn LGD formula. Rosch and Scheule (2009) have formulated and estimated the model for a time series of aggregate portfolio level default and recovery rates. However, in practice probabilities of default and recovery rates need to be estimated on the level of each individual receivable and the estimation of the model parameters should mimic this practice. Our contribution is to formulate and estimate a cross-sectional, i.e. exposure level model. The model allows incorporating receivable-specific as well as systematic explanatory variables. The estimation of the model extends the estimation of the classical Probit or Logit model incorporating the information on recovery rates on defaulted receivables (see also Bade et al., 2011). Analogously to Witzany (2001) we prefer the Bayesian MCMC estimation procedure, rather than empirically difficult likelihood maximization. The advantage of the Bayesian estimation approach is that we are also able to estimate the latent systematic factors and analyze consistently significance and various confidence intervals of the estimated parameters. The theoretical model and the estimation procedure are outlined in the following section. Section 3 then presents the empirical results based on the Moody's default and recovery database. Finally, we will compare the current Basel III capital requirements and the theoretical requirements under the proposed model and make a conclusion.

2 Two Systematic-Factor Default and Recovery Rate Model

We will focus on the model proposed by Rosch, Scheule (2009). The model captures the event of default on exposure level driven by a set of known idiosyncratic, known systematic, and an unknown (latent) systematic factor. The recovery rates (and the complementary loss given default rates) are defined similarly with a different latent systematic factor. Specifically, the event of default of a receivable *i* is driven by the time *t* normally distributed "score"

$$S_{it} = -\gamma_0 - \boldsymbol{\gamma}\mathbf{z}_{i,t-1} + \omega F_t + \sqrt{1-\omega^2}\,\xi_{it}^D \tag{2}$$

where $\mathbf{z}_{i,t-1}$ is a vector of explanatory factors known at time $t-1$, F_t a systematic normally distributed latent factor and ξ_{it} a specific (independent and normally distributed) latent factor, both representing the change between times *t* and $t-1$, and ω is the asset correlation.

Following the classical argument used for the regulatory Vasicek's formula (see, e.g., Witzany, 2010b), the expected default rate conditional on a systematic factor value $F_t = f_t$ is

$$CDR(f_t) = \Phi\left(\frac{\gamma_0 + \boldsymbol{\gamma}\mathbf{z}_{i,t-1} - \omega f_t}{\sqrt{1-\omega^2}} \right). \tag{3}$$

In order to express the default rate conditional on the explanatory factors, but not on the latent systematic factor, we need the lemma that is formulated and proved in Appendix 1.

According to the lemma, in case of the conditional default rate (3) we can integrate:

$$\int_{-\infty}^{+\infty} \Phi\left(\frac{\gamma_0 + \boldsymbol{\gamma}\mathbf{z}_{i,t-1} - \omega f_t}{\sqrt{1-\omega^2}} \right) \varphi(f_t) df_t = \Phi\left(\frac{\gamma_0 + \boldsymbol{\gamma}\mathbf{z}_{i,t-1}}{\sqrt{1-\omega^2}} \Big/ \frac{1}{\sqrt{1-\omega^2}} \right) = \Phi\left(\gamma_0 + \boldsymbol{\gamma}\mathbf{z}_{i,t-1}\right).$$

Therefore, the expected default rate (PD) is $PD = \Phi\left(\gamma_0 + \boldsymbol{\gamma}\mathbf{z}_{t-1}\right)$, and so $\gamma_0 + \boldsymbol{\gamma}\mathbf{z}_{t-1} = \Phi^{-1}\left(PD\right)$, given the *PD* value estimated at time $t-1$. If the systematic factor is set to the $1-\alpha = 0.1\%$ quantile, i.e. $f_t = \Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$ then we obtain exactly the Basel II formula

$$UDR(\alpha) = \Phi\left(\frac{\Phi^{-1}(PD) + \omega\Phi^{-1}(\alpha)}{\sqrt{1-\omega^2}} \right). \tag{4}$$

Recovery Rate Model

Similarly, the recovery rate of a receivable *i* that has defaulted at time *t* is modeled as the Probit transformation $RR_{it} = \Phi(Y_{it})$ of a normally distributed variable Y_{it} decomposed into a vector of explanatory factors, latent systematic and idiosyncratic factors. Due to the effect of integration given by the lemma above and in order to keep our formulas compatible with Rosch and Scheule (2009), we consider the debtor specific driving variable in the form

$$Y_{it} = \left(\beta_0 + \boldsymbol{\beta}\mathbf{z}^R_{i,t-1} + bX_t\right)\sqrt{1+\sigma^2} + \sigma\xi^R_{it}. \tag{5}$$

Then, according to the lemma, the recovery rate conditional on the systematic factor is given by

$$\begin{aligned} RR(X_t) &= \int_{-\infty}^{+\infty} \Phi\left(\left(\beta_0 + \boldsymbol{\beta}\mathbf{z}^R_{i,t-1} + bX_t\right)\sqrt{1+\sigma^2} + \sigma\xi\right)\varphi(\xi)d\xi = \\ &= \Phi\left(\beta_0 + \boldsymbol{\beta}\mathbf{z}^R_{i,t-1} + bX_t\right), \end{aligned} \tag{6}$$

in line with Rosch and Scheule (2009). Next, integrating the systematic factor we obtain the expected recovery rate

$$ERR = \int_{-\infty}^{+\infty} \Phi\left(\beta_0 + \boldsymbol{\beta}\mathbf{z}_{i,t-1}^R + bX_t\right)\varphi(X_t)dX_t = \Phi\left(\frac{\beta_0 + \boldsymbol{\beta}\mathbf{z}_{i,t-1}^R}{\sqrt{1+b^2}}\right).$$

Therefore, if we are given the expect loss given default (*ELGD*) parameter then

$$1 - ELGD = \Phi\left(\frac{\beta_0 + \boldsymbol{\beta}\mathbf{z}_{i,t-1}^R}{\sqrt{1+b^2}}\right), \text{ and so}$$

$$\beta_0 + \boldsymbol{\beta}\mathbf{z}_{t-1} = -\Phi^{-1}(ELGD)\sqrt{1+b^2}\ . \tag{7}$$

Given a probability level, eg $\alpha = 99.9\%$, and the latent factor quantile $x_t = \Phi^{-1}(1-\alpha)$ we can express the stand-alone downturn LGD according to (6) and (7) as

$$\begin{aligned} DLGD_{\text{stand-alone}} &= 1 - \Phi\left(-\Phi^{-1}(ELGD) + b\Phi^{-1}(1-\alpha)\right) = \\ &= \Phi\left(\Phi^{-1}(ELGD) + b\Phi^{-1}(\alpha)\right). \end{aligned} \tag{8}$$

However, it would be inconsistent to multiply the stand-alone stressed PD given by (4) and the stand-alone downturn LGD given by (7). If the two systematic factors were uncorrelated then UDR should be multiplied by the expected LGD; if the systematic factors were perfectly correlated then the product of the stand-alone UDR and DLGD would be a correct choice; but if the correlation is somewhere in between then none of the approaches is correct.

Rosch and Scheule (2009) propose a correlation ρ between the two normally distributed systematic factors and define the stressed portfolio loss rate as the product of default rate and LGD conditional only on the systematic default factor $F_t = f_t$,

$$CLR(f_t) = UDR(f_t) \times DLGD(f_t) \tag{9}$$

Let us assume that the systematic factors are bivariate normal with the correlation ρ, then X_t can be written in the form $X_t = \rho F_t + \sqrt{1-\rho^2}W$ where W is a standard normal variable independent on F_t. Applying the lemma we obtain

$$DLGD(f_t) = \int_{-\infty}^{+\infty} CLGD(\rho f_t + \sqrt{1-\rho^2}\, w)\varphi(w)dx =$$
$$= 1 - \int_{-\infty}^{+\infty} \Phi(\beta_0 + \boldsymbol{\beta}\mathbf{z}_{t-1} + b\rho f_t + b\sqrt{1-\rho^2}\, w)\varphi(w)dw = \tag{10}$$
$$= 1 - \Phi\left(\left(\beta_0 + \boldsymbol{\beta}\mathbf{z}_{t-1} + b\rho f_t\right) \sqrt{\frac{1}{1+b^2\left(1-\rho^2\right)}} \right).$$

Applying (7) and setting $f_t = -\Phi^{-1}(\alpha)$ we obtain the following relatively nice analytical formula

$$DLGD(\alpha) = \Phi\left(\left(\Phi^{-1}(ELGD)\sqrt{1+b^2} + b\rho\Phi^{-1}(\alpha)\right) \sqrt{\frac{1}{1+b^2\left(1-\rho^2\right)}} \right). \tag{11}$$

Consequently, the downturn loss rate (9) is given by an analytical formula with expected PD and LGD inputs, correlation parameters ω, b, ρ, and with the probability level parameter α which could serve as an improved regulatory formula:

$$DLR(\alpha) = \Phi\left(\frac{\Phi^{-1}(PD) + \omega\Phi^{-1}(\alpha)}{\sqrt{1-\omega^2}} \right) \times$$
$$\times \Phi\left(\left(\Phi^{-1}(LGD)\sqrt{1+b^2} + b\rho\Phi^{-1}(\alpha)\right) \sqrt{\frac{1}{1+b^2\left(1-\rho^2\right)}} \right). \tag{12}$$

Generally, the two factor model downturn loss rate $DLR_{\text{2-factor}}(\alpha)$ on the probability level α should be defined as the α-quantile of the loss rate

$$CLR(F_t, X_t) = UDR(F_t) \times DLGD(X_t) \tag{13}$$

conditional on the two systematic factors F_t, X_t with a given a correlation structure. In this case, there is no analytical formula and we have to run a Monte Carlo simulation. We will empirically compare the one-factor downturn loss rate (12) and the empirical downturn loss rate based on the two-factor decomposition (13).

Estimation Methodology

In order to apply and analyze the two-factor model, we need to estimate the correlation parameters ω, b, ρ. The estimation may be based only on observed aggregate time dependent default rates and recovery rates as in Rosch, Scheule (2009). The explanatory factors can be only systematic or related to exposure pools on which the estimation is performed. However, in practice banks estimate PD and often even LGD on exposure level given all available

exposure specific information. The unexpected risk is then relative to the information contained in the known explanatory factors in line with the models (2) and (5). Therefore, the estimation based only on aggregate numbers might overestimate the unexpected risk.

Aggregate PD-RR Model

Let us firstly assume that we are given aggregate time series: $dr_t, rr_t, \mathbf{z}_{t-1}, t = 1,...,T$, where dr_t is the observed default rate over a time period (e.g. a year) t, rr_t the observed average recovery rate on the exposures that defaulted in t, and $\mathbf{z}_{t-1}$ is a vector of macroeconomic explanatory factors known at the beginning of the year t (or at the end of $t-1$). We can initially start with the same vector of potential explanatory factors for defaults and recovery rates with non-significant variables being eliminated at the end. If the observations are made on a large pool of exposures where idiosyncratic factors diversify away we can assume that the observed default rates are realizations of (3) and the observed recovery rates are realizations of (6). Therefore,

$$\begin{aligned} dr_t &= g_1(f_t) = \Phi\left(\frac{\gamma_0 + \boldsymbol{\gamma}\mathbf{z}_{t-1} - \omega f_t}{\sqrt{1-\omega^2}}\right), \\ rr_t &= g_2(x_t) = \Phi(\beta_0 + \boldsymbol{\beta}\mathbf{z}_{t-1} + b x_t), \end{aligned} \tag{14}$$

where $\langle f_t, x_t \rangle \sim N\left(0; \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ are standard normal with the correlation ρ.

Given the parameters $\gamma_0, \boldsymbol{\gamma}, \beta_0, \boldsymbol{\beta}, \omega, b$ and the explanatory variables $\mathbf{z}_{t-1}$ the systematic factors f_t, x_t can be obtained by inverting the equations (14). The likelihood of the pair of observations $\langle dr_t, rr_t \rangle$ conditional on the parameters and the explanatory variables then equals to the bivariate normal density $\varphi_2\left(\langle f_t, x_t \rangle; \rho\right)$ divided by the Jacobian of the transformation $\langle g_1, g_2 \rangle$. Since the observations, conditional on the explanatory variables, are assumed to be independent the unknown parameters $\gamma_0, \boldsymbol{\gamma}, \beta_0, \boldsymbol{\beta}, \omega, b, \rho$ can be estimated maximizing the total likelihood function:

$$L = \prod_t \frac{\varphi_2\left(\langle f_t, x_t \rangle; \rho\right)}{\left|g_1'(f_t) g_2'(x_t)\right|},$$

or rather its logarithm $\ln L$, where

$$f_t = g_1^{-1}(dr_t) = \frac{\gamma_0 + \gamma z_{t-1} - \sqrt{1-\omega^2}\,\Phi^{-1}(dr_t)}{\omega},$$

$$x_t = \frac{\Phi^{-1}(rr_t) - \beta_0 - \beta z_{t-1}}{b},$$

$$\left|g_1'(f_t)\right| = \frac{\omega}{\sqrt{1-\omega^2}}\varphi\left(\Phi^{-1}(dr_t)\right), \text{ and}$$

$$\left|g_2'(x_t)\right| = b\varphi\left(\Phi^{-1}(rr_t)\right).$$

The terms $\varphi\left(\Phi^{-1}(dr_t)\right)$, $\varphi\left(\Phi^{-1}(1-lgd_t)\right)$ do not depend on the parameters to be estimated and can be taken out during the maximization. In order to make the estimation computationally efficient we maximize the log-likelihood $\ln L$ where the independent terms are taken out:

$$\widetilde{LL} = \sum_t \left(\frac{-(f_t^2 - 2\rho f_t x_t + x_t^2)}{2(1-\rho^2)} - 0.5\ln(1-\rho^2) - \ln\omega + 0.5\ln(1-\omega^2) - \ln b \right). \qquad (15)$$

The maximization can be performed numerically and the parameter variance can be obtained from the inverse Fisher information matrix (Greene, 2003). Alternatively, we may apply the Bayesian Markov Chain Monte Carlo (MCMC) simulation, specifically the Metropolis-Hastings random walk algorithm (see Appendix 2). The advantage of the approach is that we obtain a full Bayesian distribution, and hence confidence intervals, of the estimated parameters.

Cross-Sectional PD-RR Model

As explained above, it is preferable to estimate the parameters given historical records of individual defaults and recovery rates in case of default. Moreover, the cross-sectional model differs from the aggregate model by allowing individual debtor information in the default and recovery rate drivers (2) and (5). Let us assume that we are given a set of observations of exposures *i* with $d_{it} \in \{0,1\}$ indicating default at the end of the period *t*, recovery rate $rr_{it} \in (0,1)$ observed if $d_{it} = 1$, and $\mathbf{z}_{i,t-1}$ the vector (systematic and exposure specific) of explanatory factors known at the beginning of the period *t*. Following the approach of Bade et al. (2011) we set up the following likelihood function conditional on the unknown systematic factors f_t and x_t:

$$L_t(f_t, x_t) = \prod_{i=1}^{n_t} CPD_{it}^{d_{it}} (1 - CPD_{it})^{1-d_{it}} h(rr_{it})^{d_{it}} \text{ , where}$$

$$CPD_{it} = \Phi\left(\frac{\gamma_0 + \boldsymbol{\gamma}\mathbf{z}_{i,t-1} - \omega f_t}{\sqrt{1-\omega^2}} \right) \tag{16}$$

and $h(rr_{it})$ is the probability density according to the model specification (5). That is $rr_{it} = \Phi(Y_{it}) = f(\xi_{it}^R)$, $\xi_{it}^R \sim N(0,1)$, and so

$$h(rr_{it}) = \frac{\varphi(\xi_{it}^R)}{\varphi(\Phi^{-1}(rr_{it}))\sigma} \text{ , where}$$

$$\xi_{it}^R = \frac{\Phi^{-1}(rr_{it}) - \left(\beta_0 + \boldsymbol{\beta}\mathbf{z}_{i,t-1}^R + bX_t\right)\sqrt{1+\sigma^2}}{\sigma}. \tag{17}$$

In order to estimate the parameters by MLE we firstly need to integrate out the latent systematic factors from the total conditional likelihood, i.e.

$$L = \prod_{t=1}^{T} \iint_{f_t y_t} L_t(f_t, \rho f_t + \sqrt{1-\rho^2}\, y_t) df_t dy_t \text{ where}$$

$x_t = \rho f_t + \sqrt{1-\rho^2}\, y_t$ is decomposed into two normally distributed components with $y_t \sim N(0,1)$ independent on f_t.

Bade et al. (2011) have implemented this approach in a one-factor model numerically integrating out the systematic factor, but the estimation becomes numerically even more difficult and less stable given the two systematic factors requiring numerical double integration. Therefore, we will prefer again the Bayesian MCMC algorithm outlined in Appendix 2 where the latent factors are sampled along with the unknown model parameters. To make the estimation efficient we work with the following modified log-likelihood function where we eliminate those components that do not depend on the variables estimated:

$$\widetilde{LL} = \sum_t \sum_i \left(d_{it} \ln CPD_{it} + (1 - d_{it}) \ln(1 - CPD_{it}) - 2 \ln \xi_{it} - \ln \sigma \right). \tag{18}$$

In fact, estimating the latent factors f_t and x_t we can work only with the part depending on *t*, etc.

3 Empirical Study

In order to estimate the parameters of the outlined aggregate and cross-sectional PD-LGD model we use the Moody's Corporate Default Risk Service (DRS) database which contains for almost 36 000 corporate and sovereign entities and more than 525 000 debts. The data

spans from 1970 to 2011 and contains the Moody's rating history, default and recovery information, debt enhancement, issuer industry and other basic information. Since the database contains only 1058 sovereigns where just 42 defaults (out of 2258 total observed defaults) were observed, we perform the study without separating the corporate and sovereign entities. Lagged U.S. GDP growth[2] from 1969 to 2011 will be used as a global macroeconomic variable in line with Casseli et al. (2008) or Belyaev et al. (2012).

Aggregate Model

Regarding the aggregate model, Figure 1 shows the annual default rates calculated as the number of issuers that defaulted during a year divided by the number of all rated issuers at the beginning of the year according to DRS. Moreover, it shows the average recovery rate of all exposures that defaulted during that year. The two series are apparently visually negatively correlated, in particular since the mid-eighties. Complementarily, we expect the default rate – LGD correlation to be positive. In order to deal with autocorrelation and external dependencies in the aggregate model (14) we use the lagged default rates (Φ^{-1}transformed) and US GDP growth as the default rate explanatory variables. The lagged average recovery rate (Φ^{-1}transformed) and the US GDP growth are used as the recovery rate explanatory variables. Since the number of observed defaults in years 1971-1981 has been very low (less than 10 per year, only 1 in 1979 and 2 in 1981) we can hardly assume that the specific recovery risk has been diversified away taking the annual averages. Therefore, we have used only the observations spanning the years 1982-2011 where the number of defaults is at least 10 per year. Table **1** shows the estimation results based on 5000 MCMC iterations where we dropped the first 1000 iterations. Figure 2 and Figure 3 indicate a relatively good convergence of the estimation procedure for the two parameters b and ρ. The results in Table **1** show that the default rate series is (not surprisingly) strongly auto-correlated (the coefficient γ_1), the recovery rate series surprisingly does not show a significant autocorrelation (the coefficient β_1), and that the dependence of the default rates and LGDs on the US GDP growth is weak (coefficients γ_2 and β_2). The estimated default and LGD correlations (systematic factors' loading coefficients ω and b) turn out to be relatively large (21.8% and 31.3%), positive, and significant with a low estimation error. The default – recovery rate correlation ρ mean estimate is as expected positive 49%, however, with a larger estimation error (15.8%), and being significant on the 5% probability level.

[2] U.S. Department of Commerce Bureau of Economic Analysis (www.bea.gov)

Figure 1: Annual default rate (left axis) of all rated issuers and average recovery rates (right axis) of all defaulted issues in the Moody's DRS database 1970-2011

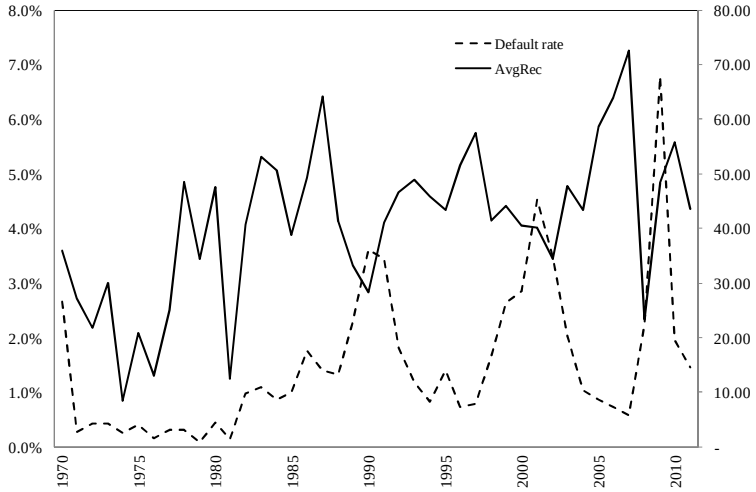


Table 1: Estimation results based on 5000 MCMC iterations with 1000 burnout period

	γ_0	γ_1	γ_2	β_0	β_1	β_2	ω	b	ρ
Mean	-0.7818	0.6237	1.3048	-0.1118	-0.0894	0.5842	0.2179	0.3128	0.4901
Std	0.3229	0.1565	1.5744	0.0968	0.2148	2.3036	0.0285	0.0446	0.1594
$\mathbf{q_{5\%}}$	-1.2552	0.3840	-1.3591	-0.2642	-0.4308	-4.0723	0.1749	0.2506	0.1961
$\mathbf{q_{95\%}}$	-0.2186	0.8984	3.9674	0.0557	0.2631	3.6651	0.2694	0.3935	0.7281

Figure 2: MCMC iterations (left chart) and the Bayesian distribution (right chart) of the correlation parameter ρ

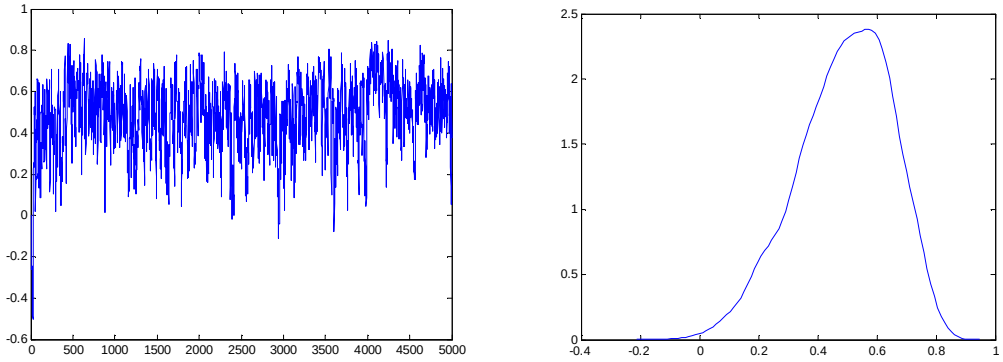
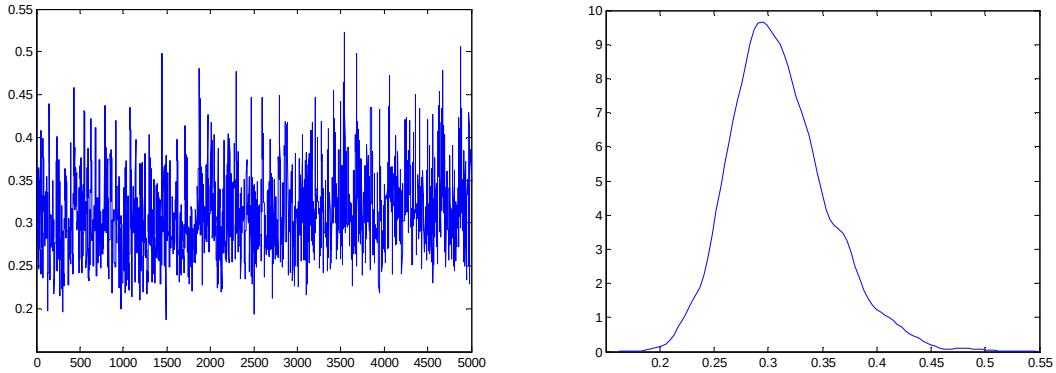


Figure 3: MCMC iterations (left chart) and the Bayesian distribution (right chart) of the correlation parameter (LGD systematic factor loading) b



Cross-sectional Model

In order to estimate the cross-sectional model (16) we had to select a random subsample from the set of all observations that can be obtained from the DRS database. By an observation we mean an exposure rated at the beginning of a year, with a default indicator at the end of the year, and an observed LGD value in case the default took place. Since there are more than 380 000 exposures, and each can be observed for several years, there are over 1 million of possible observations. However, the numerical MCMC procedure (implemented in Matlab) based on the likelihood function (18) takes hours already with 5 000 exposures in spite of the proposed efficiency improvements.

Similarly to default logistic regression function development practice, we have selected a random subsample of 5000 cases with 2500 defaults and 2500 non-defaults. The defaulted cases are given a larger weight in order to capture better the information on realized recovery rates. Regarding explanatory factors, we have used debt specific information given by the rating and seniority at the beginning of the observation period, lagged average default rate, lagged average recovery rate, and the lagged US GDP growth that was used again as a global macroeconomic indicator. The categorical rating and seniority information were translated into numerical variables using average default rates and realized LGDs based on the DRS database (see Figure 4) and transformed in both cases by the inverse normal cumulative distribution function Φ^{-1}.

The estimations results shown in Table 2 are based on 5 000 MCMC iterations. As indicated by Figure 5, in this case it was necessary to discard the first 2000 iterations. The results show a strong explanatory power of the rating and seniority variables (coefficients γ_1 and β_1), a

weaker explanatory power of the lagged default rate (γ_2), non-significant lagged recovery rate similarly to the aggregate model (β_2). The default rate sensitivity to the US GDP growth (γ_3) is significant with the negative sign (as expected), while the recovery rate sensitivity to the US GDP growth (β_3) is weakly significant positive (again as expected). The mean estimate of the default and recovery rate systematic factor loadings ($\hat{\omega} = 0.266$ and $\hat{b} = 0.286$) come out relatively close to the estimates from the aggregate model and with a low estimation error. The estimation error of $\hat{\rho} = 0.62$ is larger (0.12), but it is significant on the 5% confidence level. The difference between the aggregate model and the cross-sectional model (Table 2) is more pronounced in case of the default – recovery correlation. This can be generally explained by the fact that the models use different explanatory factors and that the parameters ω, b, ρ characterize the residual variability and correlation not explained by these explanatory factors. As explained in the introduction, the cross-sectional model corresponds better to the banking practice where account level PDs and LGDs are estimated. The MCMC procedure also estimates, as a by-product, the default (F_t) and recovery rate (X_t) systematic factors. The sampled mean values and 90% confidence intervals are shown in Figure 6. For example, the significant drop of both factors during 2008 corresponds to an unexpected increase in default rates and an unexpected decline of the recovery rates.

Our results, based on the cross-sectional model, are consistent with the results of Rosch, Scheule (2009) where the Moody's data ending in 2007 were also used, but only on the aggregate level, and separately for various rating and seniority pools.

Figure 4: Default rates conditional on rating (left chart) and average recovery rates conditional on seniority (right chart) according to the DRS database

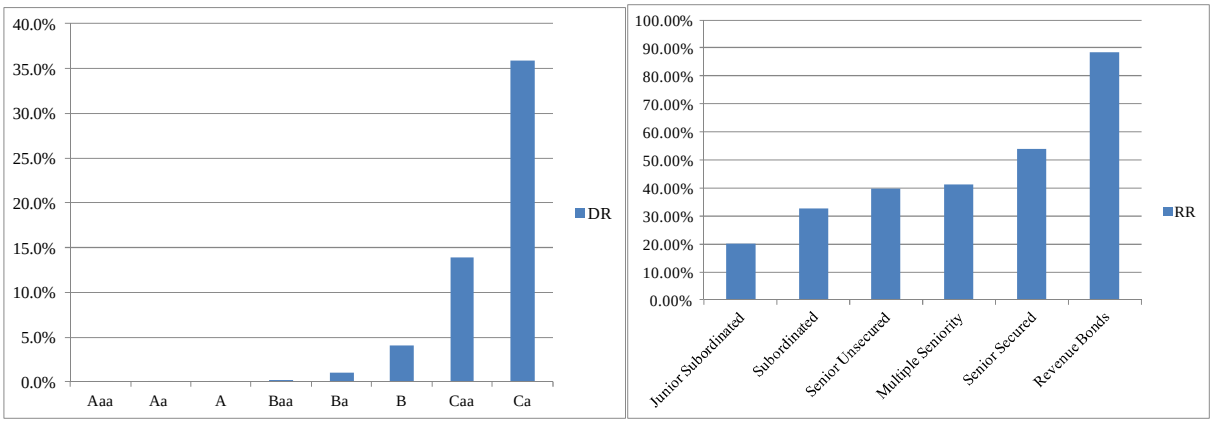


Table 2: Cross-sectional model estimation results based on 5000 MCMC iterations and 2000 burnout period

	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	ω	b	ρ	σ
Mean	0.4788	1.5328	0.2916	-2.479	2.1520	1.1711	-0.123	0.6852	0.2661	0.2864	0.6199	0.9787
Std	0.1000	0.0728	0.1339	0.7351	0.2550	0.0396	0.1090	0.6471	0.0435	0.0412	0.1213	0.0136
$\mathbf{q_{5\%}}$	0.3359	1.4111	0.0652	-3.453	1.7420	1.2307	-0.351	-0.571	0.2059	0.2240	0.4035	0.9561
$\mathbf{q_{95\%}}$	0.6464	1.6514	0.5027	-1.077	2.4993	1.1038	0.0285	1.5355	0.3403	0.3639	0.7932	1.0012

Figure 5: MCMC iterations (left chart) and the Bayesian distribution (right chart) of the correlation parameter ρ (cross-sectional model)

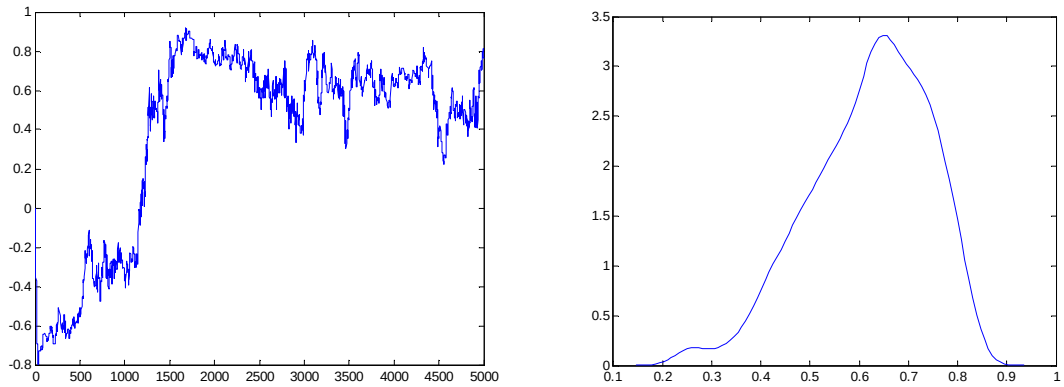
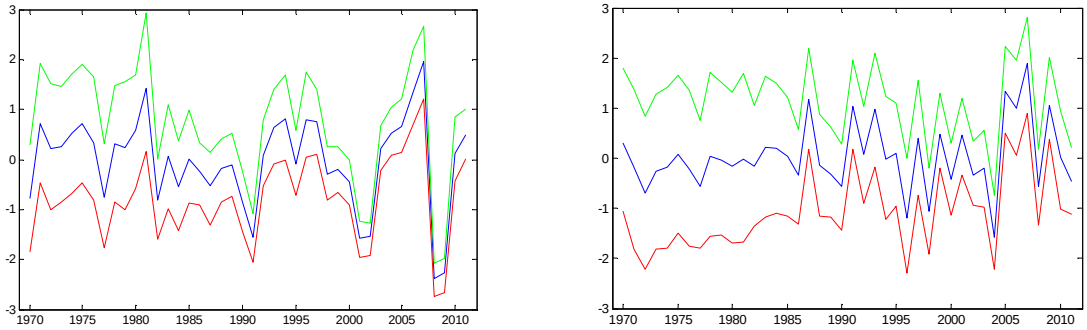


Figure 6: Estimated default (left chart) and recovery rate (right chart) systematic factors and their confidence intervals (90%)



Unexpected loss Estimation

We are going to compare four different approaches to unexpected loss estimation on the portfolio of equally weighted investments in issues from the DRS database that were assigned

a valid rating as of 1.1.2012. There are 928 issues satisfying this condition and our portfolio value is 928 million USD, assuming that 1 million USD has been invested into each of those issues. For each issue we use the expected probability of default given by the 2011 rating and the expected recovery rate conditional on the seniority of the issue as key inputs of our models:

- **Four-factor model** will be the model where the event of default and the recovery rate in case of default are driven by the variables (2) and (5). In order to simulate a scenario we have to sample the two correlated systematic factors common for the portfolio, and then the two independent idiosyncratic factors for every issue *i* in the portfolio. The portfolio loss in a scenario is calculated as $\sum_{i=1}^{N} EAD_i \times Def_i \times LGD_i$ where $Def_i \in \{0,1\}$ is the default indicator determined by the simulated default driver variable, LGD_i the simulated loss given default, and N = 928 the number of issues in the portfolio. To estimate the desired loss quantiles we need to run the Monte Carlo simulation sufficiently many times.
- **Two-factor model** will be based on the equations (3) and (6), i.e. in this case only the two correlated systematic variables are sampled and the loss conditional on those factors is calculated as $\sum_{i=1}^{N} EAD_i \times UDR_i(F_t) \times DLGD_i(X_t)$. This model implicitly assumes that the specific risk is diversified away and takes into account only the risk of the two systematic variables. Again, the loss distribution is sampled by the Monte Carlo simulation.
- **Reduced two-factor model** calculation is based on the formula (12). In this case, no simulation is needed. Given a probability level α the unexpected loss is directly calculated as $UL(\alpha) = \sum_{i=1}^{N} EAD_i \times UDR_i(\alpha) \times DLGD_i(\alpha)$ where $DLGD_i(\alpha)$ is given by (11). The model is called reduced two-factor because it is based on the two factor model, but the unexpected loss is conditional only on the appropriate quantile of the first (default-related) systematic factor.
- **Single factor-model** is the current Basel II model based on (4) and on a vague downturn LGD concept. In order to make the LGD input more precise, we will stress the parameter by the stand-alone formula (8) given a probability level α_1. In this case

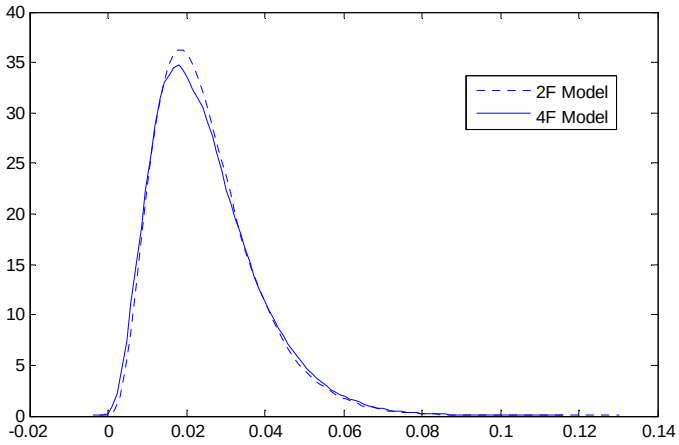
the unexpected loss is again calculated without any Monte Carlo simulation directly by the formula $UL(\alpha) = \sum_{i=1}^{N} EAD_i \times UDR_i(\alpha) \times DLGD_i(\alpha_1)$.

The parameters used in the computation are the mean estimates from Table 2, ie $\omega = 0.27$, $b = 0.29$, $\rho = 0.62$, and $\sigma = 0.98$. The Monte Carlo simulation has been run 100 000 times. The estimated unexpected loss rates (as a percentage of total exposure) according to the four models and with different α, α_1 values are given in Table 3. The results could be compared with the expected loss rate of 2.44%. The average expected PD on the portfolio is 3.91% and the average expected recovery rate is 39%. The results are in line with our expectations: the unexpected loss according to the four-factor model is larger than in the two-factor model since the former takes into account the idiosyncratic risk which does not diversify perfectly even in the large testing portfolio (see Figure 7). The results of the reduced two-factor model on different probability levels are only slightly below the two-factor model. Therefore, the reduced two-factor model provides a very good approximation of unexpected loss quantiles. The unexpected loss according to the (Basel II) one-factor model is dramatically lower if we use the concept of expected or median LGD ($\alpha_1 = 50\%$). The last three rows in Table 3 show the results for different LGD stressing levels. The interesting conclusion is that LGD must be stressed at least on the 95% level (with 97.5% being more or less optimal in this case) in order to get comparable values.

Table 3: Unexpected loss rates (as a percentage of total EAD) estimated by the four models and for different α, α_1 values

	α=95%	α=99%	α=99.9%
Four-Factor Model	5.02%	6.59%	8.72%
Two-Factor Model	4.91%	6.46%	8.42%
Reduced Two-Factor Model	4.82%	6.29%	8.27%
One-Factor Model (α_1=50%)	4.16%	5.15%	6.42%
α_1=90%	5.01%	6.21%	7.75%
α_1=95%	5.22%	6.48%	8.09%
α_1=97.5%	5.39%	6.69%	8.36%

Figure 7: Loss distributions in the four-factor and two-factor models



4 Conclusion

This study compares the current regulatory one-factor approach to unexpected loss estimation and the two-factor model proposed by Rosch, Scheule (2009). The advantage of the model is that it captures consistently the recovery rate variation and its correlation with the rate of default. We have proposed two approaches how to estimate the model parameters: based on aggregate default rate and recovery rate time series and a cross-sectional approach based on exposure level data. In both cases our estimation procedure uses the MCMC Bayesian approach. The empirical results (based on the Moody's DRS database) confirm not only significant variability of the recovery rate but also a significant correlation over 50% between the rate of default and the recovery rates in the context of the model. Our empirical comparison has shown that the reduced two-factor model analytical formula proposed by Rosch, Scheule (2009) performs well compared to simulated results (based on our estimated parameter values). In contrast, the performance of the regulatory formula is poor and heavily depends on the discretionary conservatism in LGD stressing. In our case, approximately 97.5% probability level LGD stressing would be needed, but this level could differ for different datasets or products depending on the default and recovery rate correlations. Our main conclusion is that the reduced two-factor analytical formula works well and could feasible replace the current regulatory formula with regulatory parameters based on the presented or similar empirical studies.

Literature

[1] **Acharya, Viral, V., S. Bharath and A. Srinivasan (2007).** Does Industry-wide Distress Affect Defaulted Firms? – Evidence from Creditor Recoveries, Journal of Financial Economics 85(3):787–821.

[2] **Altman E., Resti A., Sironi A. (2004).** Default Recovery Rates in Credit Risk Modelling: A Review of the Literature and Empirical Evidence, Economic Notes by Banca dei Paschi di Siena SpA, vol.33, no. 2-2004, pp. 183-208

[3] **E. Altman, G. Fanjul (2004).** Defaults and Returns in the High Yield Bond Market: Analysis through 2003, NYU Salomon Center Working Paper

[4] **Bade B., Rosch D., Scheule H. (2011).** Default and Recovery Risk Dependencies in a Simple Credit Risk Model, European Financial Management, Vol. 17, No. 1, 2011, 120–144

[5] **BCBS 2005.** Basle Committee on Banking Supervision, "Guidance on Paragraph 468 of the Framework Document", Bank for International Settlements.

[6] **BCBS (2006).** Basel Committee on Banking Supervision, **"**International Convergence of Capital Measurement and Capital Standards, A Revised Framework – Comprehensive Version", Bank for International Settlements.

[7] **BCBS (2010).** Basel III: A global regulatory framework for more resilient banks and banking systems, Bank for International Settlements.

[8] **Bellotti, T. and J. Crook (2009).** "Calculating LGD for Credit Cards." QFRMC Conference on Risk Management in the Personal Financial Services Sector, January 2009.

[9] **Belyaev K., Belyaeva A., Konečný T., Seidler J., Vojtek M. (2012).** "Macroeconomic Factors as Drivers of LGD Prediction: Empirical Evidence from the Czech Republic", CNB Working Paper Series, 12/2012, p. 46

[10] **Caselli, S., S. Gatti, and F. Querci (2008).** "The Sensitivity of the Loss Given Default Rate to Systematic Risk: New Empirical Evidence on Bank Loans." Journal of Financial Services Research 34, pp. 1–34.

[11] **De Graeve, F., T. Kick, and M. Koetter (2008).** "Monetary Policy and Financial (In)stability: An Integrated Micro–Macro Approach." Journal of Financial Stability 4(3), pp. 205–231.

[12] **Frye, J. (2000a).** Collateral Damage, RISK 13(4), 91–94.

[13] **Frye, J. (2000b).** Depressing recoveries, RISK 13(11), 106–111.

[14] **Frye, J. (2003).** A false sense of security, RISK 16(8), 63–67.

[15] **Gordy, M. (2003).** A risk factor foundation for ratings based bank capital rules. Journal of Financial Intermediation, 12, pp. 199-232.

[16] **Greene W.H. (2003).** Econometric Analysis, Prentice Hall, 5th Edition, pp. 1026.

[17] **Gupton G., Gates D., and Carty L. (2000).** Bank loan losses given default, Moody's Global Credit Research, Special Comment.

[18] **Jacobson, T., R. Kindell, J. Lindé, and K. Roszbach (2011).** "Firm Default and Aggregate Fluctuations." International Finance Discussion Papers No 1029, Board of Governors of the Federal Reserve System.

[19] **Johannes M., Polson N. (2009).** MCMC Methods for Financial Econometrics , Handbook of Financial Econometrics (eds. Ait-Sahalia and L.P. Hansen), p. 1-72.

[20] **Lynch, S. M. (2007).** Introduction to Applied Bayesian Statistics and Estimation for Social Scientists, Springer, pp. 359.

[21] **Pykhtin, M. (2003).** Unexpected recovery risk, Risk, Vol 16, No 8. pp. 74-78.

[22] **Rachev S.T., Hsu J.S, Bagasheva B.S., Fabozzi F.J. (2008).** Bayesian Methods in Finance, The Frank J. Fabozzi Series, Wiley, pp. 329.

[23] **Rosch D., Scheule H. (2009).** Credit Portfolio Loss Forecasts for Economic Downturns, Financial Markets, Institutions & Instruments, V. 18, No. 1, February

[24] **Seidler, J., R. Horvath, and P. Jakubík (2009).** "Estimating Expected Loss Given Default in an Emerging Market: The Case of Czech Republic." Journal of Financial Transformation 27, pp. 103–107.

[25] **Tasche, Dirk. (2004).** The single risk factor approach to capital charges in case of correlated loss given default rates, Working paper, Deutsche Bundesbank, February

[26] **Vasicek O. (1987).** "Probability of Loss on a Loan Portfolio," KMV Working Paper, p. 4.

[27] **Witzany Jiří (2010a).** On Deficiencies and Possible Improvements of the Basel II Unexpected Loss Single-Factor Model, Czech Journal of Economics and Finance, 3, pp. 252-268

[28] **Witzany J. (2010b).** Credit Risk Management and Modeling. Praha: Nakladatelství Oeconomica, p. 212.

[29] **Witzany J (2011)**. A Two Factor Model for PD and LGD Correlation, Bull. Of the Czech Econometric Society, 18(28), pp. 1-19.

Appendix 1: A Probability Lemma

Lemma: $\int_{-\infty}^{+\infty} \Phi(a+bx)\varphi(x)dx = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right)$, where a and b are constants, Φ is the standard normal cdf, and φ is the standard normal pdf.

Proof: Since $\Phi(a+bx) = \Pr[Z < a+bx]$ where Z is standard normal, the integral on the left hand side equals to $\Pr[Z < a+bX]$ where Z and X are independent standard normal variables. Consequently,

$$\int_{-\infty}^{+\infty} \Phi(a+bx)\varphi(x)dx = \Pr[Z < a+bX] = \Pr\left[\frac{Z-bX}{\sqrt{1+b^2}} < \frac{a}{\sqrt{1+b^2}}\right] = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right)$$

since $\frac{Z-bX}{\sqrt{1+b^2}}$ is a standard normal random variable. The result can be also verified by direct integration.

Appendix 2: Bayesian MCMC Estimation Procedure

The Bayesian MCMC sampling algorithm has become a strong and frequently used tool to estimate complex models with multidimensional parameter vectors, including latent state variables. Examples are financial stochastic models with jumps, stochastic volatility processes, models with complex correlation structure, or switching-regime processes. For a more complete treatment of MCMC methods and applications we refer for example to Johannes, Polson (2009), Rachev et al. (2008), or Lynch (2007).

MCMC provides a method of sampling from multivariate densities that are not easy to sample from directly, by breaking these densities down into more manageable univariate or lower dimensional multivariate densities. To estimate a vector of unknown parameters $\Theta = (\theta_1, ..., \theta_k)$ from a given dataset, where we are able to write down the Bayesian marginal densities $p(\theta_j \mid \theta_i, i \neq j, \text{data})$ but not the multivariate density $p(\Theta \mid \text{data})$, the MCMC *Gibbs sampler* works according to the following generic procedure:

0. Assign a vector of initial values to $\Theta^0 = (\theta_1^0, ..., \theta_k^0)$ and set $j = 0$.
1. Set $j = j+1$.

2. Sample $\theta_1^j \sim p(\theta_1 \mid \theta_2^{j-1},...,\theta_k^{j-1},\text{data})$.

3. Sample $\theta_2^j \sim p(\theta_2 \mid \theta_1^j,\theta_3^{j-1},...,\theta_k^{j-1},\text{data})$.

$\vdots$

k+1. Sample $\theta_k^j \sim p(\theta_k \mid \theta_1^j,\theta_2^j,...,\theta_{k-1}^j,\text{data})$ and return to step 1.

According to the Clifford-Hammersley theorem the conditional distributions $p\left(\theta_j \mid \theta_i, i \neq j, \text{data}\right)$ fully characterize the joint distribution $p\left(\Theta \mid \text{data}\right)$ and moreover, under certain mild conditions, the Gibbs sampler distribution converges to the target joint distribution (Johannes, Polson, 2009).

The conditional probabilities are typically obtained applying the Bayes theorem to the likelihood function and a prior density, for example

$$p\left(\theta_1 \mid \theta_2^{j-1},...,\theta_k^{j-1},\text{data}\right) \propto L\left(\text{data} \mid \theta_1,\theta_2^{j-1},...,\theta_k^{j-1}\right)\cdot\text{prior}\left(\theta_1 \mid \theta_2^{j-1},...,\theta_k^{j-1}\right). \tag{19}$$

We can often use uninformative priors, ie $\text{prior}\left(\theta_i\right) \propto 1$ and assume that the parameters are independent. In order to apply the Gibbs sampler the right hand-side of the proportional relationship needs to be normalized, ie we need to be able to integrate the right-hand side with respect to θ_1 conditional on $\theta_2^{j-1},...,\theta_k^{j-1}$.

Useful Gibbs sampling distributions are univariate or multivariate normal, Inverse Gamma or Wishart, and the Beta distribution. For example, if $\mathbf{y} = \langle y_1,...,y_T \rangle$ is an observed series and assuming that iid $y_i \sim N\left(\mu,\sigma^2\right)$ with unknown parameters μ and σ then

$$\begin{aligned} p(\mu \mid \mathbf{y},\sigma) &\propto L(\mathbf{y} \mid \mu,\sigma^2)\,p(\mu) = \prod_{i=1}^{T}\varphi(y_i;\mu,\sigma) \propto \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^T \exp\left(-\frac{\sum(y_i-\mu)^2}{2\sigma^2}\right) \\ &\propto \exp\left(-\frac{T\mu^2 - 2\mu\sum y_i}{2\sigma^2}\right) \propto \varphi\left(\mu;\frac{\sum y_i}{T},\frac{\sigma}{\sqrt{T}}\right) \end{aligned} \tag{20}$$

using the uninformative prior $p(\mu) \propto 1$. Moreover,

$$\begin{aligned} p(\sigma^2 \mid \mathbf{y},\mu) &\propto L(\mathbf{y} \mid \mu,\sigma^2)\cdot p(\sigma^2) = \frac{1}{\sigma^2}\prod_{i=1}^{T}\varphi(y_i;\mu,\sigma) \\ &\propto \left(\sigma^2\right)^{-\frac{T}{2}-1} \exp\left(-\frac{\sum(y_i-\mu)^2}{2\sigma^2}\right) \propto IG\left(\sigma^2;\frac{T}{2},\frac{\sum(y_i-\mu)^2}{2}\right) \end{aligned} \tag{21}$$

using the prior $p(\sigma^2) \propto 1/\sigma^2$ equivalent to the uninformative log-variance prior $p(\log \sigma^2) \propto 1$. Hence the Bayesian distributions for μ and σ can be obtained by the Gibbs sampler iterating (20) and (21). The prior distributions are often specified in order to improve convergence but not to influence (significantly) the final results, typically a wide normal distribution conjugate prior distribution for μ and a flat inverse gamma distribution for σ^2 are used.

If the integration on the right hand side of (19) is not analytically possible (which is also our case) then the *Metropolis-Hastings algorithm* can be used. It is based on the rejection sampling algorithm. For example in step 2 the idea is firstly to sample a new proposal value of θ_1^j and then accept it or reject it (ie reset $\theta_1^j := \theta_1^{j-1}$) with appropriate probability so that, intuitively speaking, we rather move to the parameter estimates with higher corresponding likelihood values.

Specifically, step 1 is replaced with a two step procedure:

1. A. Draw θ_1^j from a proposal density $q(\theta_1 \mid \theta_1^{j-1}, \theta_2^{j-1}, ..., \theta_k^{j-1}, \text{data})$,

B. Accept θ_1^j with the probability $\alpha = \min(R,1)$, where

$$R = \frac{p\left(\theta_1^j \mid \theta_2^{j-1}, ..., \theta_k^{j-1}, \text{data}\right) q\left(\theta_1^{j-1} \mid \theta_1^j, \theta_2^{j-1}, ..., \theta_k^{j-1}, \text{data}\right)}{p\left(\theta_1^{j-1} \mid \theta_2^{j-1}, ..., \theta_k^{j-1}, \text{data}\right) q\left(\theta_1^j \mid \theta_1^{j-1}, \theta_2^{j-1}, ..., \theta_k^{j-1}, \text{data}\right)}. \tag{22}$$

In practice the step 1B is implemented by sampling a $u \sim U(0,1)$ from the uniform distribution and accepting θ_1^j if and only if $u < R$.

It is again shown (see Johannes, Polson, 2009) that under certain mild conditions the limiting distribution is the joint distribution $p(\Theta \mid \text{data})$ of the parameter vector. Note that the limiting distribution does not depend on the proposal density, or on the starting parameter values. The proposal density and the initial estimates only make the algorithm more-or-less numerically efficient.

A popular proposal density is the random walk, ie sampling by

$$\theta_1^j \sim \theta_1^{j-1} + N(0, c). \tag{23}$$

The algorithm is then called *Random Walk Metropolis-Hastings*. The proposal density is in this case symmetric, ie the probability of going from θ_1^{j-1} to θ_1^j is the same as the probability of going from θ_1^j to θ_1^{j-1} (fixing the other parameters), and so the second part of the fraction in the formula (22) for α in step 1B cancels out. Consequently, assuming non-informative prior, the acceptance or rejection is driven just by the likelihood ratio

$$R = \frac{L\left(\text{data} \mid \theta_1^j, \theta_2^{j-1}, ..., \theta_k^{j-1}\right)}{L\left(\text{data} \mid \theta_1^{j-1}, \theta_2^{j-1}, ..., \theta_k^{j-1}\right)}.$$

Another popular approach is the *Independence Sampling Metropolis-Hastings algorithm* where the proposal density $\mathrm{q}\left(\theta_1^j\right)$ does not depend on θ_1^{j-1} (given the other parameters). The acceptance probability ratio (22) is slightly simplified but the proposal densities do not cancel out. In order to achieve efficiency the shape of the proposal density q should be close to the shape of the target density p, which is known only up to a normalizing constant.

Typically, estimating complex stochastic models, we need to estimate the parameter vector with a few model parameters Θ, and a vector with a large number of state variables X (proportional to the number of observations). We know that $p(\Theta, X \mid \text{data}) \propto p(\text{data} \mid \Theta, X) \cdot p(X, \Theta)$ and so we may estimate iteratively the parameters and the state variables:

$$p(\Theta \mid X, \text{data}) \propto p(\text{data} \mid \Theta, X) \cdot p(X \mid \Theta) \cdot p(\Theta),$$
$$p(X \mid \Theta, \text{data}) \propto p(\text{data} \mid \Theta, X) \cdot p(\Theta \mid X) \cdot p(X).$$

The parameters and state variables are sampled step by step, or in blocks, often combining Gibbs and Metropolis-Hastings sampling.

In case of our aggregate model (15) we use the random walk Metropolis-Hastings with the step standard deviation *c* corresponding to the expected estimate variation. This is obtained by running the algorithm with an expertly set parameter *c* (eg 0.1 in case of the correlation parameters) and then adjusting the constant in order to achieve a reasonable acceptance rate around 40-80% (see Lynch, 2007). For the cross-sectional model we sample, in addition, the independent latent factors f_t and y_t from the bivariate normal distribution with the correlation ρ.