Forthcoming in the *Journal of Operational Risk*

Hidden Markov Regimes in Operational Loss Data: Application to the Recent Financial Crisis

Georges Dionne[1] and Samir Saissi Hassani
Canada Research Chair in Risk Management, HEC Montréal, and CIRRELT

August 29, 2016

Abstract

We propose a method to consider business cycles in the computation of capital for operational risk. We examine whether the operational loss data of American banks contain a Hidden Markov Regime switching feature from 2001 to 2010. We assume asymmetric distribution of monthly losses. Statistical tests do not reject this assumption. A high level regime is marked by very high loss values during the recent financial crisis, confirming temporal heterogeneity in the data. If this heterogeneity is not considered in risk management models, capital estimations will be biased. Banks will hold too much capital during periods of low stress and not enough capital in periods of high stress. Additional capital reaches 30% during this period of analysis when regimes are not considered.

Keywords

Hidden Markov Regime, operational risk, 2007-2009 financial crisis, skew t type 4 distribution, banks' regulatory capital.

[1] Corresponding author: Georges Dionne, HEC Montréal, 3000, Chemin de la Côte-Sainte-Catherine, room 4.454, Montreal (Qc), Canada, H3T 2A7. Phone: 514-340-6596. Email: georges.dionne@hec.ca.

1 Introduction

We propose a method to consider business cycles in the computation of capital for operational risk. The recent literature on credit risk management advocates extending the standard structural models to see how time-varying market conditions can improve credit risk predictions and solve the credit risk puzzle (Chen 2010; Bhamra et al. 2010). One interesting result recently tested by Maalaoui-Chun et al. (2014) is that credit spread levels exhibit persistence after recessions. This dynamic behavior must be taken into account when computing optimal capital for credit risk. To our knowledge, no such modeling exists in the operational risk literature. The only exception is the contribution of Allen and Bali (2007), which introduces cyclicality in operational risk measurement but does not estimate a dynamic model that endogenizes operational losses over business cycles.

Since the inception of operational risk modeling, authors have argued that the amount of reserve capital calculated is very fragile, even unstable. Ames, Schuermann and Scott (2014) demonstrate this fragility with operational loss data, particularly since the recent financial crisis that began in 2007. Neslehová, Embrechts, and Chavez-Demoulin (2006) had affirmed the risk of working with "extreme value" distributions when preliminary estimates tend to exhibit an infinite mean or variance for the data (see also Dahen et al, 2010a). These results suggest that more conventional base models should be considered to better estimate the loss distributions, particularly for computing capital with a confidence level of 99.9%.

We use the operational loss data of American banks to examine whether the data contain a Hidden Markov Regime switching feature for the 2001-2010 period. We build on the scaling model of Dahen and Dionne (2010b) and show that the operational loss data of American banks are indeed characterized by a Hidden Markov Regime switching model. The distribution of monthly losses is asymmetric, with a normal component in the low regime and a skew t type 4 component in the high regime. Statistical tests do not allow us to reject this asymmetry. We then introduce the regimes obtained in the estimation of operational losses and affirm that their presence significantly affects the distribution of losses in general. We also analyze the scaling of the data to banks of different sizes and risk exposures, and present the results of backtesting of the model in two different banks.

We mainly observe temporal heterogeneity in the data. If this heterogeneity is not considered in the risk management models, capital estimations will be biased. Levels of reserve capital will be overestimated in periods of normal losses corresponding to the low level of the regime, and underestimated in high regime periods. Banks tended to allot too much capital to operational risk when the regimes were not considered in our period of analysis.

The research on scaling models started with the article of Shih et al. (2000), which showed that institution size is the main scaling factor. Hartung (2004) developed a normalization formula based on scaling factors such as firms' revenues and quality of risk management. Na et al. (2006) proposed a model with common components and an idiosyncratic component. Dahen and Dionne (2010b) added scaling variables such as time, business line and risk type to this approach, as suggested by the Basel II regulation. They also extended the scaling methodology to the frequency of losses. In this article, we add business cycles to the model and analyze how this improvement affects the optimal regulatory capital. Other contributions on operational risk management are Cruz (2002), Lewis (2003), Wei (2007), Chapelle et al. (2008), Chernobai and Yildirim (2008), and Jarrow (2008).

In Section 2, we present the database used. Section 3 discusses identification and estimation of models of regimes. Section 4 measures the effect of regimes detected on the estimation of the distribution of operational losses, and Section 5 proposes a backtest of estimated parameters and a robustness analysis of our models. A short conclusion ends the article.

2 Data

We use the Algo OpData Quantitative Database for operational losses of $1 million and more sustained by US banks. The study period is from January 2001 to December 2010. We examine the operational losses of US Bank Holding Companies (BHC) valued at over $1 billion. The source of information on these banks is the Federal Reserve of Chicago. Statistics on the sample built from the two databases are summarized in tables 1 and 2.

Table 1 presents the size distribution of banks with $1 billion or more in assets that were exposed to operational losses of $1 million or more during the study period. We note a major increase in the mean size of banks during this period; maximum size has also grown significantly. Table 2 shows that the largest banks accumulated the largest losses. Table 3 presents the Event Types and Business Lines codes subject to operational losses, as defined by the Basel regulation.

Table 1: Number of BHC banks per year and their assets

	Assets (in billions $)				
Year	Median	Mean	Max	Sd	Number
2001	2.1	19.7	944.3	82.3	356
2002	2.1	19.5	1,097.2	84.8	378
2003	2.0	20.3	1,264.0	93.0	408
2004	2.0	25.4	1,484.1	122.1	421
2005	2.0	24.4	1,547.8	121.9	445
2006	2.1	26.0	1,884.3	140.5	461
2007	2.1	28.9	2,358.3	168.1	460
2008	2.0	28.5	2,251.5	182.5	470
2009	2.1	33.8	2,323.4	190.6	472
2010	2.1	34.7	2,370.6	198.3	458

Note: Sd is standard deviation.

Table 2: Operational losses of BHC banks with bank assets in deciles

Asset deciles (in billions $)	Loss (in millions $)					
	Min	Max	Median	Mean	Sd	Number
2,022.7 to 2,370.6	1.0	8,045.3	26.3	265.9	1,129.5	51
1,509.6 to 2,022.7	1.0	8,400.0	14.0	268.3	1,207.5	49
1,228.3 to 1,509.6	1.0	2,580.0	7.5	94.5	357.8	53

799.3 to 1,228.3	1.0	3,782.3	24.0	199.8	610.7	48
521.9 to 799.3	1.0	8,400.0	7.4	218.9	1,156.4	53
1,247.1 to 521.9	1.1	210.2	7.2	17.0	31.1	50
98.1 to 247.1	1.0	663.0	6.0	45.3	115.4	51
33.7 to 98.1	1.0	775.0	10.2	55.2	152.8	51
8.31 to 33.7	1.1	691.2	8.6	32.2	98.6	51
0.96 to 8.31	1.0	65.0	4.3	9.9	14.5	51
All	1.0	8,400.0	8.6	120.1	680.7	508

Note: Sd is standard deviation.

Table 3: Nomenclature of Event Types and Business Lines codes

Variables	Codes
Event Types	
Clients, Products and Business practices	CliPBP
Business disruption and system failures	BusDSF
Damage to physical assets	DamPA
Employment practices and workplace safety	EmpWS
External fraud	EF
Internal fraud	IF
Execution, Delivery and Process Management	ExeDPM
Business Lines	
Retail brokerage	RBr
Payment and settlement	PayS
Trading and sales	TraS
Commercial banking	ComB
Retail banking	RBn
Agency services	AgnS
Corporate finance	CorF
Asset management	AssM

3 Identification of regimes

We assume that there are regimes in operational loss data. Figures 1 and 2 support this assertion.[2] The hatched area in figure 1 identifies the dot-com recession in 2001 and the recent recession corresponding to the financial crisis that began in 2007. The number of operational losses increased significantly during the last financial crisis, which did not occur during the 2001 recession. We observe another spike in the number of losses in 2010, one year after the recession ended. The losses in 2010 may be explained by delays linked to lawsuits. Indeed, several banks were sued after the financial crisis. Figure 2 presents similar evolutions in loss volatility.

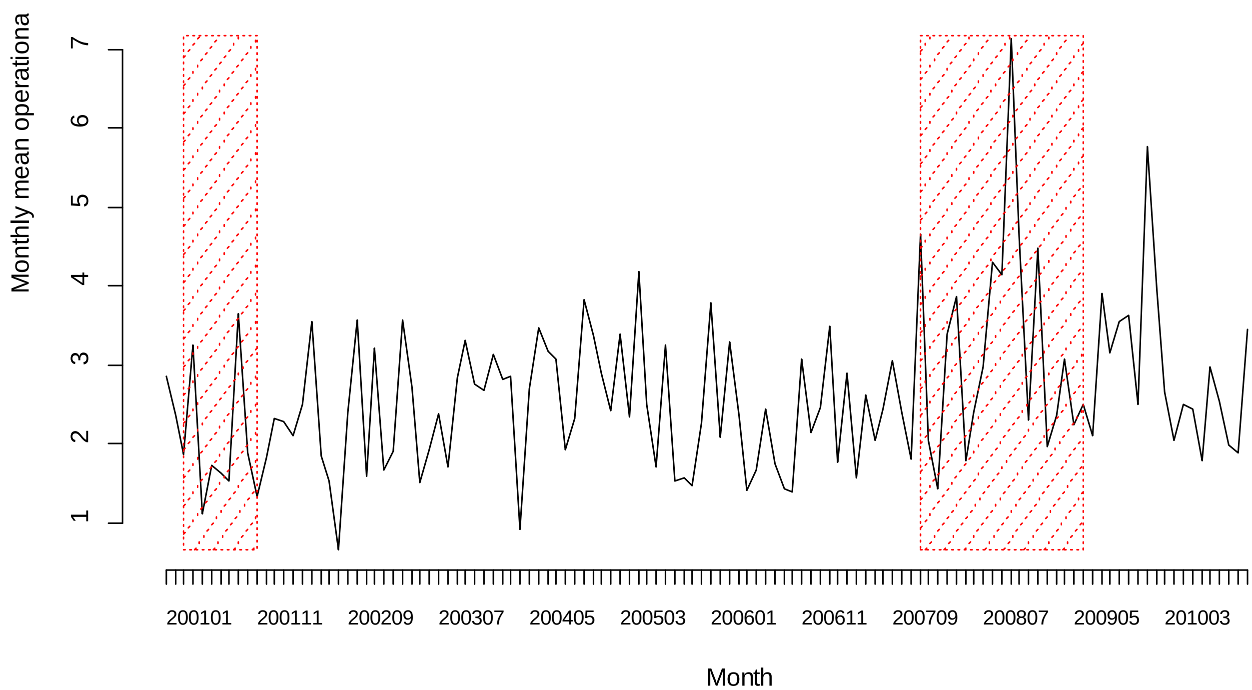

Figure 1: Changes in monthly mean operational losses

[2] HMM does not yield distinct patterns of the level and volatility in the time series (see Maalaoui et al, 2014, for more details). Figures 1 and 2 illustrate too few differences to justify an approach that would separate mean regimes from volatility regimes.

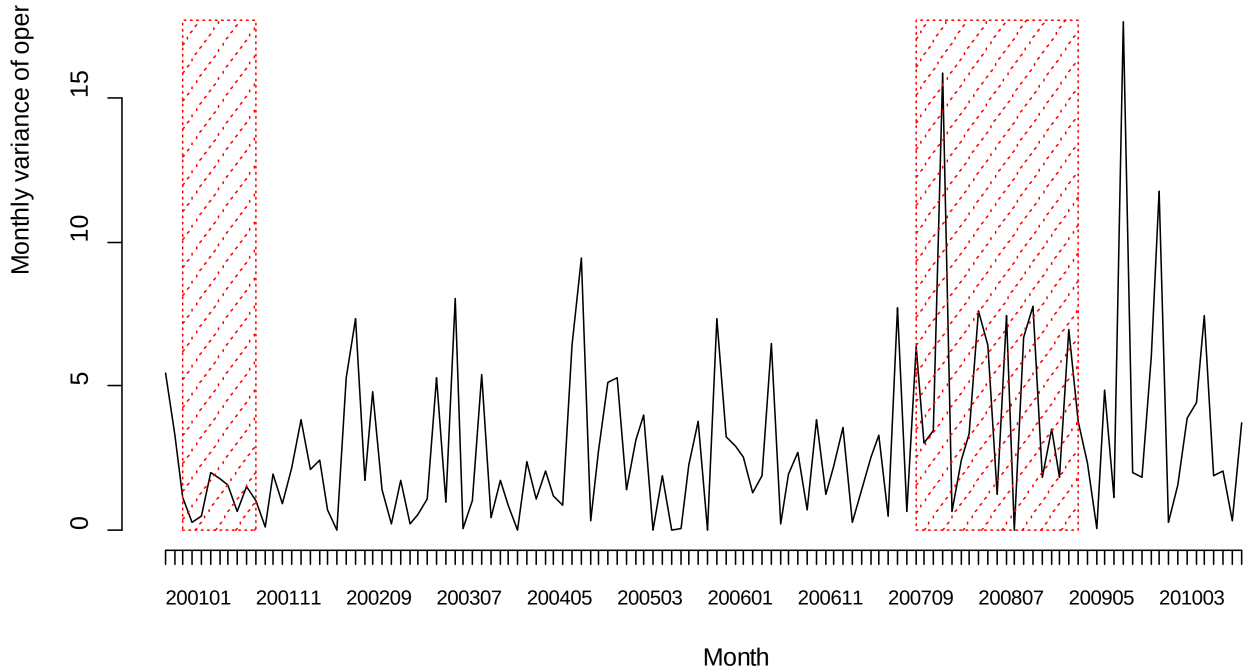

Figure 2: Changes in monthly variance of operational losses

3.1 Markov Switching Regimes

3.1.1 Markov Switching Model

We suppose that the data[3] under study represent a system that possesses *n* possible distinct states. At any given moment, the system may be in either state. For a given state, the system can move to another state or remain in place. Given that states are not observable, the model is called a Hidden Markov Model, or HMM. For our data, the objective is to identify and characterize "high loss" periods (state 2, for example) and separate them from "normal loss" periods (state 1) endogenously. We inject information on loss severity and frequency that comes uniquely from the data, such that the model will show the unobservable underlying dynamics. We also analyze a three-state application in the robustness section of the paper.

[3] Another possibility is to model the regime switching with individual severities instead of aggregate severities. In this paper, we estimate the HMM with industry data because we want to measure the effect of industry regimes on banks' losses.

3.1.2 Estimation of the HMM with the Baum-Welch method

To develop the estimation, we follow Zucchini and MacDonald (2009), Mitra and Date (2010), and Visser and Speekenbrink (2010). We now define the necessary notations. The variables are indexed by time $t \in \{1,2,...,T-1,T\}$. Observations are noted as x_t. The sequence of observations from $t=a$ to $t=b$ is noted as $x_{a:b} = x_a, x_{a+1},...,x_{b-1}, x_b \, (a,b = 1 \text{ to } T)$. The variable s_t represents the state where the system is situated at time $t, s_t \in \{1,...,n\}$. Suppose that *n* states exist. Similarly, $s_{a:b} = s_a, s_{a+1},...,s_{b-1}, s_b$ is the sequence of states of the system in the time interval *a* to *b*. The estimation will give a vector of the parameters θ. The model is supposed to depend on the covariates noted as z_t. According to Mitra and Date (2010), an HMM is well defined when the parameters $\{A,B,\pi\}$ are known: *A* is the transition matrix $n \times n$ whose elements are written as $a_{ij} = Pr(s_{t+1} = j | s_t = i, z_t, \theta)$, *B* is a diagonal matrix whose elements $b_i(x_t) = Pr(x_t | s_t, z_t, \theta)$ are written according to the densities that describe x_t when the system is in the state *i* = 1 to *n*, and π is a row vector $(1 \times n)$ of the probabilities related to each state at $t=1$, $\pi_i = Pr(s_1 = i | z_t, \theta)$, $\pi = (\pi_1,...,\pi_i,...,\pi_n)$. To simplify the presentation, we examine the case of two states $(n=2)$, f_1 being the density function of a normal law for the low-loss regime (state 1), f_2 being the density function of the skew t-distribution type 4 representing the high-loss regime (state 2).[4] The choice of this mixture of distributions will be justified later when we present the estimation results. For now, note that $B_t = \begin{bmatrix} f_1(x_t) & 0 \\ 0 & f_2(x_t) \end{bmatrix}$ such that:

$$f_1(x_t | \mu_1, \sigma_1^2) = \frac{1}{\sigma_1 \sqrt{2\pi}} exp\left[\frac{(x_t - \mu_1)^2}{2\sigma_1^2} \right] \qquad (3.1)$$

where $\sigma_1 > 0$ and $\mu_1 \in \mathbb{R}$.

The skew t type 4 distribution, noted as ST4, is defined as in Rigby et al. (2014):

[4] For the first application of the skew-t distribution for operational risk, see Allen and Bali (2007).

$$f_2\left(x_t \middle| \mu_2,\sigma_2,\nu,\tau\right)=\frac{c}{\sigma_2}\left\{\left[1+\frac{\left(x_t-\mu_2\right)^2}{\nu\sigma_2^2}\right]^{-(\nu+1)/2} I\left(x_t<\mu_2\right)+\left[1+\frac{\left(x_t-\mu_2\right)^2}{\tau\sigma_2^2}\right]^{-(\tau+1)/2} I\left(x_t\geq\mu_2\right)\right\} \quad (3.2)$$

where

$$\sigma_2,\nu,\tau>0,\mu_2\in\mathbb{R},c=2\left[\nu^{1/2}B\left(1/2,\nu/2\right)+\tau^{1/2}B\left(1/2,\tau/2\right)\right]^{-1}.$$

B is the beta function. $B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)$ where Γ is the gamma function. The ST4 distribution is a shape-spliced distribution. It comprises two Student's t-distributions with different shape parameters ν and τ: the left part with ν and the right part with τ, which lets us consider a longer tail, which is important for extreme operational losses. When $\nu=\tau$, the ST4 distribution becomes the Student's t-distribution.

It is well known that operational losses are not normally distributed and are often skewed, with one heavy tail. Splicing of distributions is used to introduce skewness and kurtosis into a symmetric distribution family. The advantage of the skew t-distribution is its flexibility. It can approach normality in some circumstances and can depart from normality for data with long tails. The ST4 will be compared with the skew normal distribution, which is obtained by using different scale parameters.

Concerning the matrix, $A_t=\begin{bmatrix}\left(1-a_{12}\right) & a_{12}\\ \left(1-a_{22}\right) & a_{22}\end{bmatrix}$, the elements a_{ij} will be modeled according to a constant and m independent covariates $z_t^k, k\in\{1,...,m\}$. We denote the vector of covariates as $z_t=\left(1,z_t^1,...,z_t^m\right)$ and posit that:

$$a_{ij}=logistic\left(\eta_{ij}z_t^T\right) \quad (3.3)$$

where $logistic(\bullet)$ is the logistic function $\left(\frac{exp(\bullet)}{1+exp(\bullet)}\right),\eta_{ij}=\left(\eta_{ij,0},...,\eta_{ij,k}...,\eta_{ij,m}\right),\eta_{ij,0}$ is a constant, $\eta_{ij,k}$ is the coefficient to estimate for the k^{th} covariate z_t^k relative to the conditional probability

a_{ij}, and z_t^T denotes the transpose of z_t. We will estimate the initial distribution π as a vector of constants. We can separate the model parameter vector θ into three independent parts. Accordingly, we rewrite $\theta = (\theta_0, \theta_1, \theta_2)$ where θ_0, θ_1 and θ_2 are, respectively, the parameters to estimate for the initial distribution π, the parameters related to matrix *A* and those concerning matrix *B* representing conditional densities f_i. We now write the probability of jointly observing the sequence of observations $x_{1:T}$ and that of the states of the system $s_{1:T}$:

$$Pr\left(x_{1:T}, s_{1:T} \middle| z_{1:t}, \theta\right) = Pr\left(s_1 \middle| z_1, \theta_0\right) \prod_{t=2}^{T} Pr\left(s_t \middle| s_{t-1}, z_{t-1}, \theta_1\right) \prod_{t=1}^{T} Pr\left(x_t \middle| s_t, z_t, \theta_2\right) \quad (3.4)$$

$$log\,Pr\left(x_{1:T}, s_{1:T} \middle| z_{1:t}, \theta\right) = log\,Pr\left(s_1 \middle| z_1, \theta_0\right) + \sum_{t=1}^{T-1} log\,Pr\left(s_{t+1} \middle| s_t, z_t, \theta_1\right) + \sum_{t=1}^{T} log\,Pr\left(x_t \middle| s_t, z_t, \theta_2\right). \quad (3.5)$$

Given that equation 3.5 is formed of a sum of three independent quantities, the maximum probability can be estimated for each of the vectors of parameters θ_0, θ_1 and θ_2 separately. In addition, if we consider that the initial distribution is independent from z_1, we can estimate the *n* probabilities of the vector $\pi = (\pi_1, ..., \pi_n)$ as constants $(\theta_0 = (\pi_1, ..., \pi_n))$.

Note that the probability function to maximize depends on the sequence $s_{1:T}$ which is not observable. Our objective is to extract it from the sequence $x_{1:T}$. One technical solution is to use the EM (Expectation Maximization) concept, which is better known as the Baum-Welch algorithm. The method is presented in Dionne and Saissi (2016).

Concretely, we construct the sequence $x_{1:T}$ from monthly mean losses (in log). We use a mixture where the first "normal" state will be modeled by a normal distribution and the second state of the high regime (HR) will be represented by a skew t-distribution type 4 (ST4). We want to capture the asymmetry and thickness of the distribution tail during this state. We also use the number of losses per quarter. To do so, we create a variable called $L_i, i = 1,2,$ as a natural logarithm of the number of losses announced during the three previous months. The idea is to capture whether the number of losses announced affects the intensity of transitions of the regime from one level to the other. In short, we use four distributions as follows:

$$
\begin{aligned}
x_t \left| s_t = 1 \sim N\left(\mu_1, \sigma_1\right)\right. \\
x_t \left| s_t = 2 \sim ST4\left(\mu_2, \sigma_2, \nu, \tau\right)\right. \\
a_{12} = logistic\left(\eta_{12,0} + \eta_{12,1} L_1\right) \\
a_{22} = logistic\left(\eta_{22,0} + \eta_{22,1} L_2\right).
\end{aligned}
\tag{3.6}
$$

Lastly:

$$\theta_0 = \left(\pi_1, \ldots, \pi_n\right), \theta_1 = \left(\eta_{12,0}, \eta_{12,1}, \eta_{22,0}, \eta_{22,1}\right) \text{ and } \theta_2 = \left(\mu_1, \sigma_1, \mu_2, \sigma_2, \nu, \tau\right).$$

3.1.3 Results and discussion

We estimated four models, each of which has a density in each state. Operational risk distributions can be heavy-tailed, particularly in the high regime. It is then important to find a distribution that does not underestimate the tail quantiles and the regulated capital. We used three criteria for selecting the final distribution: goodness of fit, analysis of pseudo-residuals and estimation of skewness and kurtosis.

We first started with a four-parameter distribution in the high regime, the GB2 distribution often used in the literature. The GB2 includes, as particular cases, many distributions with heavy tailed distributions (Cummins et al. 1990; McDonald et al. 1995). Due to space limitations, the detailed estimation results are presented in Dionne and Saissi (2016). Since our data are in logarithms, the exponential GB2 (EGB2) was analyzed. The EGB2 has a very similar likelihood function and pseudo-residuals results to the selected distribution below but underestimates the kurtosis and skewness. As Carillo et al. (2012) note, in some circumstances it may be impossible to obtain the desired tail behavior with a single parametric function. Splicing two different parametric densities may yield better results for the tail estimation.

In Table 4 we present two models with spliced distributions in the high regime to better fit the tail. We begin with the parameters of the distributions that we use in each state. In Model 1, we assume a normal distribution in each state $(N+N)$, where the two density functions are as in (3.1) but with different parameters μ and σ in each state. The results in Table 4 indicate that they are indeed different. We observe a higher mean and a higher standard deviation in the high-loss regime (high period). Model 1 does not let us consider asymmetry in the high-loss regime.

In Model 2 $(N+SN)$, we integrate such asymmetry by first considering the skew normal distribution (SN) in state 2 while keeping the normal distribution in state 1. A skew normal can be represented by the following density function (Fernandez et al., 1995):

$$f_2\left(x_t \middle| \mu_2,\sigma_2,\gamma\right)=c\left\{\exp\left[-\frac{1}{2}\left(\frac{x_t-\mu_2}{\sigma_2}\right)^2\gamma^2\right]I\left(x_t<\mu_2\right)+\exp\left[-\frac{1}{2}\left(\frac{x_t-\mu_2}{\sigma_2}\right)^2\frac{1}{\gamma^2}\right]I\left(x_t\geq\mu_2\right)\right\}$$

where $c=\sqrt{2/\pi}\,\dfrac{\gamma}{\sigma_2\left(1+\gamma^2\right)}$, $\mu_2\in\mathbb{R}$, $\sigma_2>0$, $\gamma>0$ and I is an indicator variable. γ is an asymmetry measure. When $\gamma=1$, the SN distribution becomes the normal distribution.

As indicated in Model 2 of Table 4, the consideration of the SN distribution in state 2 does not improve the estimation according to the AIC criteria. Moreover, the log (γ) parameter is not significantly different from zero, meaning that $\gamma=1$, rejecting the asymmetry of the normal distribution in state 2. In state 1, the estimated parameters of the normal distribution are very similar to those obtained in Model 1.

We now consider Model 3, which corresponds to the normal distribution in state 1 and the ST4 distribution in state 2 (N+ST4). The normal distribution, which models phases of low losses, has a mean of 2.4172 and a standard deviation of 0.7653. The two corresponding parameters are very significant and are similar to those in Model 1 and Model 2. Regarding the skew-t type 4 distribution, its mean is estimated at 3.7872, whereas its standard deviation can be fixed to 1 (its log can be considered statistically null because it is non-significant). In the high regime, we still have a significant and simultaneous increase in the mean and in the standard deviation compared with state 1. In addition, the asymmetry of the skew-t type 4 is confirmed by the $\log(\nu)$ coefficient, significant at 10%, whereas the $\log(\tau)$ coefficient is non-significant. We will validate these distributions below by performing a robustness analysis of our statistical results. According to the AIC and log likelihood criteria, Model 3 outperforms the two other models but is similar to the EGB2.

Table 4: Estimation of the Hidden Markov Model

	Model 1 $N+N$		Model 2 $N+SN$		Model 3 $N+ST4$	
	Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
Probability of transition to high regime						
	Intercept	0.2879 (0.7392)	Intercept	0.3318 (0.7648)	Intercept	0.9772 (0.8161)
	L_1	-1.4853*** (0.2001)	L_1	-1.5468*** (0.2766)	L_1	-1.7371*** (0.2798)
Probability of staying in high regime						
	Intercept	-25.3101*** (4.6082)	Intercept	-28.1742*** (5.2129)	Intercept	-25.7285*** (4.5707)
	L_2	11.5681*** (2.5745)	L_2	12.9137*** (3.1253)	L_2	11.7434*** (2.4739)
Response distributions						
Low regime	Normal law		Normal law		Normal law	
	μ_1	2.4277*** (0.4366)	μ_1	2.4570*** (0.4546)	μ_1	2.4172*** (0.5876)
	σ_1	0.7685*** (0.2214)	σ_1	0.7979*** (0.2494)	σ_1	0.7653*** (0.2006)
High regime	Normal law		SN		ST4	
	μ_2	4.0294*** (6.5251)	μ_2	3.3991** (1.5123)	μ_2	3.7872*** (0.5449)
	σ_2	1.2968*** (0.1683)	σ_2	1.0207*** (0.2370)	$\log(\sigma_2)$	-0.0415 (0.2546)
			$\log(\gamma)$	0.5401 (0.7609)	$\log(\nu)$	2.7734* (1.4299)
					$\log(\tau)$	0.9492 (0.8007)
Log likelihood		-152.566	-151.863		-148.838	
AIC criteria		323.132	323.726		319.677	
Number of observations		120	120		120	

Notes: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. The log likelihood of the EGB2 is -148.785 and the AIC criteria is 319.570.

The estimation of Table 4 gives a value of $\nu = exp(2.7734) = 16.013$ and $\tau = exp(0.9492) = 2.584$, which measures a very large thickness of ST4 distribution tails. Nonetheless, given that the estimation of $\log(\tau)$ in Table 4 is non-significant, $\log(\tau)$ can be considered null, therefore $\tau = 1$. The right distribution tail would be thicker in this sense.

We now discuss the stages of the transition probability in Table 4. The coefficient of the variable L_1 is significantly negative in the three models. Note that the number of losses is historically limited to between 7 and 20 per quarter. In Table 5, we present the estimated skewness and kurtosis obtained from the results. We observe that the ST4 distribution provides the better estimation of both measures. Figure 3 shows that the tail of the ST4 remains higher than those of the other distributions to the right of the log (loss) equal to 8 on the X axis.

Table 5: Skewness and kurtosis analysis with different distributions in the high regime

	Skewness	Kurtosis
Data (log)	1.192	5.974
N	0.000	3.000
SN	0.653	3.319
ST4	1.119	6.127
EGB2	1.008	5.093

Figure 4 shows the Markov switching states detected with Model 3. Three facts emerge from the figure. First, there was almost no reaction to the recession of 2001 (2001-03 to 2001-11), and only a few fluctuations in probability transition around 2003-2005. In contrast, there is indeed a high regime detected during the recession starting in 2007 (2007-12 to 2009-06), with a first impetus lasting one month in December 2007, followed by two other variations. The first lasts five months, from July to November 2008 inclusively, and the second lasts six months, from August 2009 until January 2010 inclusively. The second impetus happens after the end of the recession.

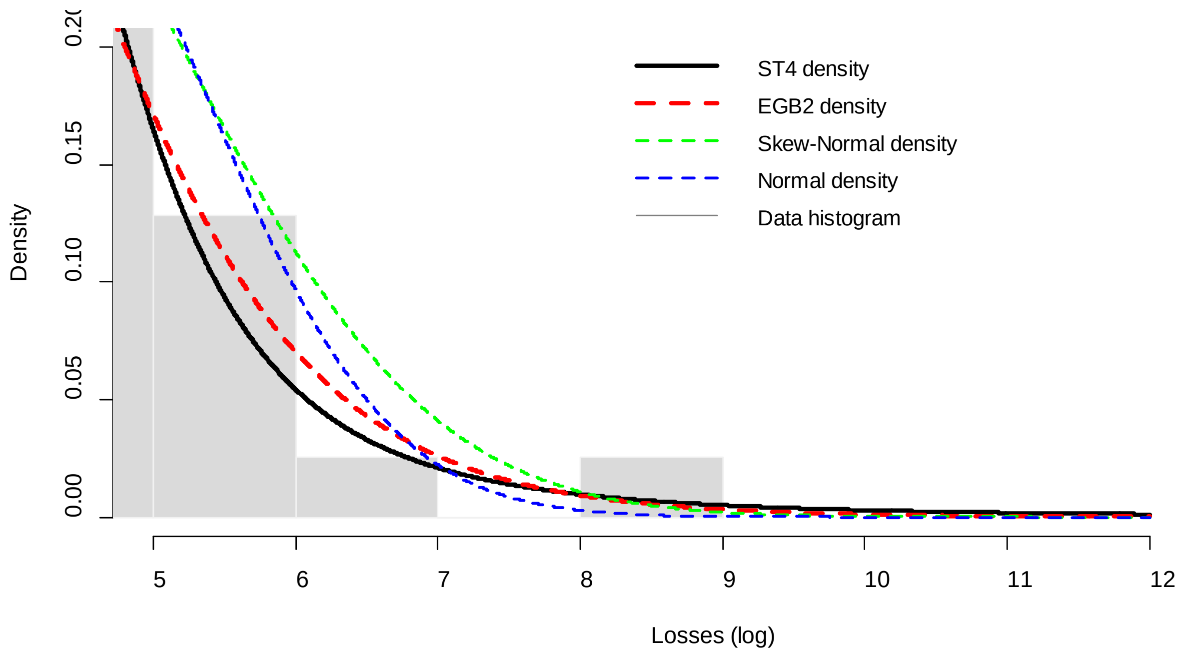

Figure 3: Data histogram and right tail density of log of losses for different distributions

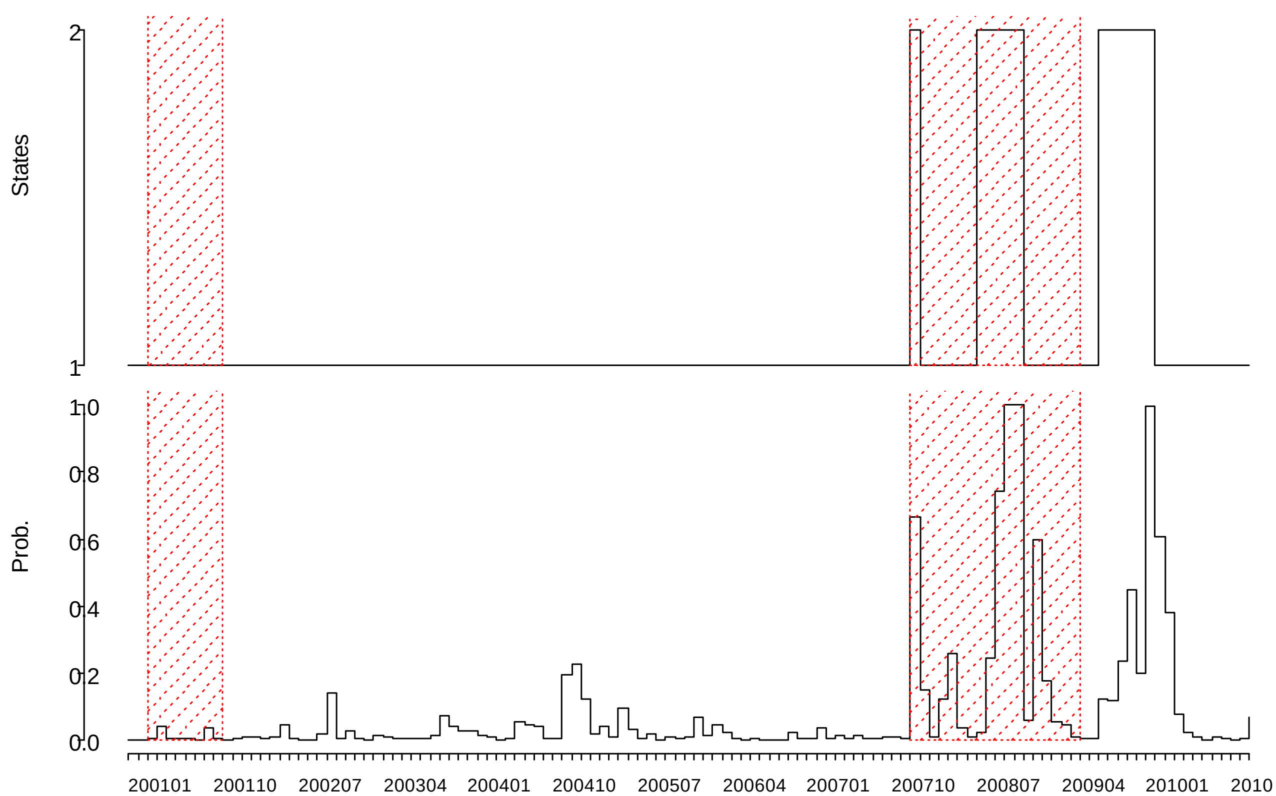

Figure 4 Markov Regimes detected from January 2001 to December 2010

We now analyze precisely what happened for the two variations.[5] To do so, we take the individual losses at the largest amounts, which represent at least 80% of the total lost during each period examined. We obtained information on the circumstances of these losses by gathering comments inserted in the loss database, which includes the Bloomberg and SEC (U.S. Securities and Exchange Commission) sites.

As reported in Table 6, there were two losses of $8.4 billion each for the first variation. This amount is an all-time record for operational losses of BHC banks. The two losses represent over 81% of the $20.6 billion lost during this first variation from July to November 2008. Both cases pertain to problems related to subprime loans. In addition, both banks agreed to settle the class-action suits without waiting for a decision from the courts. There was thus no gap between the time the problems were observed and the date the losses were reported. This is not the case for most of the large losses in the period of the second variation, from August 2009 to January 2010. Table 7 shows six major losses for this period, which account for more than 80% of the total losses. These losses were subject to varying delays due to lawsuits. Consequently, the second peak fundamentally consists of a series of problems that arose during the financial crisis. The gap in time between the two variations seems to stem uniquely from legal procedures.

Table 6: Summary of losses of BHC banks from July 2008 to November 2008

	Bank	Loss	EventType	BusLine	Date	% Loss
1	Wachovia Bank	8.4 billion	CliPBP	RBn	2008-07-21	40.73
2	CFC – Bank of America	8.4 billion	CliPBP	RBn	2008-10-06	40.73
	Other (< 80%)	3.4 billion	30 losses			
	All	20.6 billion	32 losses			

[5] A more detailed discussion is presented in Dionne and Saissi (2016).

Table 7: Summary of losses of main BHC banks from August 2009 to February 2010

	Bank	Loss	EventType	BusLine	Date	% Loss
1	Citibank N.A.	840 million	ExeDPM	TraS	2010-01-19	20.77
2	Discover Financial Service	775 million	CliPBP	RBn	2010-02-12	19.16
3	JP Morgan Securities Inc.	722 million	CliPBP	CorF	2009-11-04	17.85
4	State Street Global Advis	663 million	CliPBP	AssM	2010-02-04	16.39
5	Merrill Lynch and Company	150 million	CliPBP	CorF	2010-02-22	3.71
6	Bank of America Corporation	142 million	EF	ComB	2009-09-21	3.51
	Other (< 80%)	753 million	21 losses			
	All	4.05 billion	27 losses			

3.1.4 Specification Test of the Hidden Markov Model

Below, we statistically test the validity of the HMM specification for our data, by following Zucchini and MacDonald (2009). In general, if a random variable y follows a law $\Im$ whose cumulative function is F, the random variable defined by $u = F(y)$ must follow a uniform law $U(0,1)$. By noting as Φ the cumulative function of the normal law, we should then have:

$$y \sim \Im \Rightarrow u = F(y) \sim U(0,1) \Rightarrow \Phi^{-1}(F(y)) \sim N(0,1).$$

The variable obtained by $z = \Phi^{-1}(F(y))$ is called a pseudo-residual. If the specification $\Im$ suits the data, the pseudo-residuals should follow a normal distribution.

In our case, the vector of the pseudo-residuals of our Hidden Markov Model (Model 3) can be calculated with $z_t = \Phi^{-1}\left[Pr\left(y \le y_t \middle| y_{t-1}, \ldots, y_1\right)\right] \Rightarrow z_t \sim N(0,1)$.

Figure 5 shows the following results. The distribution of the monthly losses (in log) is asymmetric (upper panel). The middle panel reproduces the monthly losses (in log) along with the estimated *N + ST4* distribution where the normal component (*N*) is on the left-hand side. The skew t type 4 (*ST4*) component is situated to the right of the mean to take this asymmetry into account. The distribution of pseudo-residuals looks quite close to normal (bottom panel). This is confirmed by the statistical tests in Table 8. We use three tests—Kolmogorov-Smirnov, Anderson-Darling and

Shapiro-Wilk, to ensure the normal distribution of the pseudo-residuals. For comparison purposes, Table 8 shows the result of the same tests done on the series of monthly mean losses (monthly losses, in log). Because of high asymmetry, the three tests reject normality at 5% for this series of losses, as expected.

As for our model, the Anderson-Darling test gives a p-value of 0.0682 (Pseudo-residuals). This test does not reject normality even if the p-value is not far from 5%. Moreover, the Kolmogorov-Smirnov and Shapiro-Wilk tests do not allow us to reject the normality of these pseudo-residuals, with p-values of 0.1540 and 0.1560 respectively. Therefore, despite a problem of a fat-tailed distribution demonstrated by the Anderson-Darling test, we can validate our Hidden Markov specification given the two other tests and especially the Shapiro-Wilk test, which measures the global probability relative to a normal distribution. Similar results were obtained with the EGB2.

Table 8: Statistical tests

	Test	Monthly losses		Pseudo-residuals	
		Statistic	p-value	Statistic	p-value
1	Kolmogorov-Smirnov	0.1035	0.0039	0.0718	0.1540
2	Anderson-Darling	0.3101	0.0020	0.6940	0.0682
3	Shapiro-Wilk	0.9331	0.0000	0.9831	0.1560

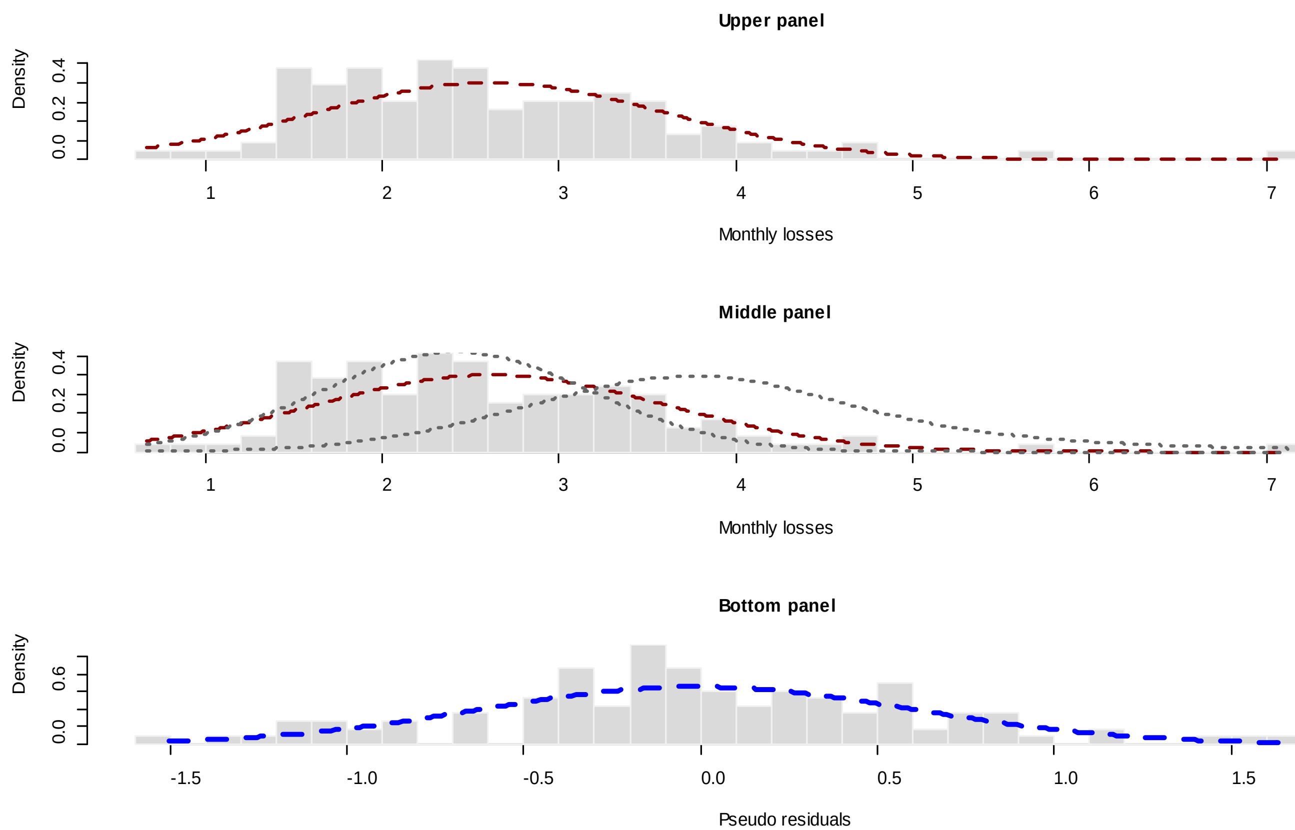

Figure 5: Histograms of monthly losses and pseudo-residuals

4 Effects of regimes detected

We start with the loss estimation model of Dahen and Dionne (2010b):

$$log\left(Loss_{it}\right) = \alpha + \beta\, log\left(Assets_{it}\right) + \sum_j \lambda_j BL_{ijt} + \sum_k \delta_k ET_{ikt} + \sum_t \gamma_t T_t + \varepsilon_{it}\ . \tag{4.1}$$

The dependent variable is *log* $(Loss_{it})$ of bank *i* in period *t*. The independent variables are *log* $(Assets_{it})$, category variables Business Lines, BL_{ijt}, where *j* is for business line *j*, and category variables EventTypes, ET_{ikt}, where *k* is for event type *k*. The fixed time effects are dummy variables for years, T_t, and ε_{it} are independent error terms assumed normally distributed $N\left(0,\sigma^2\right)$.

We add the variable of the HMM regime in model (4.2) and its cross-loadings (interaction) with Business Lines and Event Types in (4.3).

$$log\left(Loss_{it}\right)=\alpha+\beta\, log\left(Assets_{it}\right)+\sum_{j}\lambda_{i}BL_{ijt}+\sum_{k}\delta_{k}ET_{ikt}+\sum_{t}\gamma_{t}T_{t}+\xi RHMM+\varepsilon_{it}^{1}. \qquad (4.2)$$

where $RHMM$ is a dummy variable equal to one in the High HMM regime and equal to zero otherwise.

$$\begin{aligned}log\left(Loss_{it}\right)=\alpha+\beta\, log\left(Assets_{it}\right)+\sum_{j}\lambda_{i}BL_{ijt}+\sum_{k}\delta_{k}ET_{ikt}+\sum_{t}\gamma_{t}T_{t}+\xi RHMM\\+\sum_{j}\lambda_{j}^{1}BL_{ijt}\times RHMM+\sum_{k}\delta_{k}^{1}ET_{ikt}\times RHMM+\varepsilon_{ij}^{11}.\end{aligned} \qquad (4.3)$$

where λ_{j}^{1} and δ_{k}^{1} are the coefficients for the interactions between explanatory variables and regimes in place. ε_{it}^{1} and ε_{it}^{11} are the normally distributed error terms.

The regression results are presented in Table 9. Model 4.1 is the reference model. To simplify the presentation of the estimates, we do not report the coefficients of the year fixed effects coefficients (γ_t), because they are not pertinent to the discussion. "Yes" indicates their presence in the table. All standard deviations and p-values are robust to the presence of heteroskedasticity and clustering in the sense of White (1980).

Table 9: Effect of regimes detected on *log (Loss)*

Variable	(4.1) Reference model	(4.2) Adding HMM regime	(4.3) Adding HMM regime and interaction
Intercept	-0.297 (0.433)	-0.260 (0.446)	-0.160 (0.436)
Log(Assets)	0.139*** (0.037)	0.139*** (0.038)	0.126*** (0.036)
High HMM Regime		0.977*** (0.331)	1.538* (0.791)
Paymt and Settlmnt	1.261*** (0.438)	1.199*** (0.438)	1.196** (0.466)
Trading and Sales	1.104*** (0.290)	1.026*** (0.304)	0.906** (0.372)
Comm. Banking	1.182*** (0.167)	1.117*** (0.164)	1.159*** (0.172)

Retail Banking	0.930*** (0.207)	0.867*** (0.207)	0.827*** (0.171)
Agency Services	1.223*** (0.413)	1.161*** (0.435)	1.532*** (0.443)
Corp. Finance	2.056*** (0.237)	2.063*** (0.250)	1.999*** (0.294)
Asset Mngmt	1.358*** (0.274)	1.321*** (0.254)	1.307*** (0.283)
Bus.Disrup. syst.Fail.	-1.080 (0.687)	-0.926 (0.569)	-0.878 (0.630)
Damage Phy.Assets	-0.086 (1.925)	-0.044 (1.923)	0.047 (1.953)
Employ.Prac.Wrkplac.Saf.	-0.676*** (0.252)	-0.622** (0.254)	-0.476** (0.224)
External Fraud	-0.502*** (0.157)	-0.489*** (0.161)	-0.433** (0.170)
Internal Fraud	-0.593*** (0.227)	-0.524** (0.226)	-0.304 (0.211)
Exer. Deliv. Proc. Mnmt	-0.214 (0.228)	-0.217 (0.230)	-0.130 (0.256)
High Regime × Employ.Prac.Wrkplac.Saf.			-2.321*** (0.513)
High Regime × External Fraud			0.120 (1.088)
High Regime × Internal Fraud			-3.314*** (0.547)
High Regime × Exec. Deliv. Proc. Mngmt			0.115 (1.228)
High Regime × Paymt and Settlmnt			-0.561 (1.584)
High Regime × Trading and Sales			0.317 (1.248)
High Regime × Comm. Banking			-1.511 (1.266)
High Regime × Retail Banking			0.401 (1.075)
High Regime × Agency Services			-4.491*** (1.114)
High Regime × Corp. Finance			0.645 (1.565)
High Regime × Asset Mngmt			-0.249 (0.963)

Year fixed effects	yes	yes	yes
Adj. R^2	0.170	0.186	0.223
AIC	1993.52	1985.23	1978.04
Log Likelihood p-value Chi2	-971.8	-966.6 0.001 (4.2 vs 4.1)	-952.0 0.002 (4.3 vs 4.2)
Num. obs.	508	508	508

Note: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. Clients, Products and Business Practices, Retail Brokerage, and Year 2001 are the omitted categories for Event Types, Business Lines, and Years respectively.

The coefficient of the variable *log* $(Assets_{it})$ is very significant, which is consistent with this type of model, and confirms the results of Dahen and Dionne (2010b). The BL and ET coefficients tend to keep the same magnitude in all regressions. The coefficient of the high regime variable is very significant at 1% in model 4.2 but less significant in model 4.3, at 10%. In contrast, three interaction variables are significant at 1%. The presence of year fixed effects does not prevent the regimes from being significant. This suggests that the regimes detected cannot be explained by time. Comparison of the adjusted R^2 of the models shows an advantage in injecting the high regime variable in 4.2 or interaction in 4.3. The AIC statistic and the Log Likelihood ratio test also confirm the superiority of model 4.3. That being said, we must perform backtesting on these models to evaluate their validity and calculate the reserve capital. Note that the loss database did not contain observations concerning BusDSF or DamPA where the Markov regime is high. This is why the coefficients corresponding to the cross-loadings are not presented in column 4.3.

We must measure the effect of the regime levels on the loss frequencies to perform the backtest. We build the model around the zero-inflated negative binomial as in Dahen and Dionne (2010b). Let Y be a random variable that follows a negative binomial law with average λ and the dispersion parameter δ. If f_{NB} is the probability density function of this law, then the probability that Y is equal to a value k is written as:

$$Pr\left(Y=k|\lambda,\delta\right)=f_{NB}\left(k,\lambda,\delta\right)=\frac{\Gamma\left(k+1/\delta\right)}{k!\Gamma\left(1/\delta\right)}\left[\frac{1}{1+\delta\lambda}\right]^{\frac{1}{\delta}}\left[\frac{\delta\lambda}{1+\delta\lambda}\right]^{k} \quad (4.4)$$

where $k = 0,1,2,...$, and $\Gamma(\cdot)$ designates the conventional gamma function. Note that $\delta > 0$ when there is overdispersion. Further, the negative binomial converges toward a Poisson law when $\delta \to 0$ (Dionne, 1992). When there are reasons to think that there are too many 0 values relative to a negative binomial, we should envision a model with a zero-inflated negative binomial law. Let Y_{ij} be a variable representing the number of losses sustained by bank *i* for the year *j*. If Y_{ij} follows a zero-inflated negative binomial law, we can write:

$$Pr\left(Y_{ij} = k\right) = \begin{cases} q_{ij} + \left(1 - q_{ij}\right) f_{NB}\left(0, \lambda_{ij}, \delta\right) & k = 0 \\ \left(1 - q_{ij}\right) f_{NB}\left(k, \lambda_{ij}, \delta\right) & k = 1,2,... \end{cases} \tag{4.5}$$

where λ_{ij} is the mean and δ is the dispersion parameter of the basic negative binomial law, and q_{ij} represents the proportion of zeros that would be too high relative to a negative binomial law. Conditionally on the explanatory variables chosen, λ_{ij} and q_{ij} are estimated using the two following equations:

$$\begin{aligned} log\left(\lambda_{ij}\right) = \zeta_0 + \zeta_1 log\left(Assets_{ij}\right) + \zeta_2 RHMM + \zeta_3 GDP_j \\ + \zeta_4 Bank\text{-}Cap_{ij} + \zeta_5 Mean\text{-}Salary_{ij} \end{aligned} \tag{4.6}$$

$$log\left(\frac{q_{ij}}{1 - q_{ij}}\right) = \xi_0 + \xi_1 log\left(Assets_{ij}\right) + \xi_2 RHMM + \xi_3 GDP_j + \xi_4 Mean\text{-}Salary_{ij}. \tag{4.7}$$

The last formula is equivalent to modeling q_{ij} using the logistic distribution. The variable *log (Assets$_{it}$)* is the total assets of bank$_i$ (in log in period *j*) and the variable *RHMM* is a dummy variable equal to one in the High *HMM* regime. Mean-Salary$_{ij}$ is the mean salary paid in the bank *i* in period *j*, Bank-Cap$_{ij}$ is the bank *i* capitalization in period *j* and GDP$_j$ is US Gross Domestic Product during the period *j*.

The estimates are presented in Table 10. The dependent variable is the number of annual losses. In Model 1, we present the benchmark model to compare the effect of adding regimes. We used 4,329 observations from January 2001 to December 2010, as documented in Table 1. We want to measure the effect of the HMM (high) regime in both the counting and zero parts. The idea is that during high regimes, we want to see whether inflated zeros are more numerous or not.

Model 2 adds this dimension in both parts. The High Regime coefficient is negative and significant at 10% in the count, and very significantly positive for zeros. Apparently, during high levels of the Markov regime, losses would be less numerous because the zeros come more from the inflation of the zeros (outside the negative binomial). The variable GDP is also very significant to explain excess zeros. To measure whether deflation of zeros provides statistical value, we compare this deflation model with the base Model 1. Given that the models are embedded, we performed the likelihood ratio test, whose results appear in the same table. The likelihood ratio test of Model 2 versus 1 is conclusive, with a statistic of 46.53 and a p-value of almost 0. Model 2 using the Markov regime seems to provide more information than the reference Model 1 given the substantial decrease in the AIC criterion and the result of the likelihood ratio test. A final comment concerns the values of the log delta dispersion parameter of the negative binomial model. Starting with a value of 2.097 in Model 1, we reach 1.085 for Model 2, which is a clear improvement in the specification in the sense that there is less unobserved heterogeneity in Model 2. We can thus proceed to the backtesting of the model.

Table 10: Effect of regimes on frequencies

	Model 1 Reference model	Model 2 Adding HMM regime
Count model		
Intercept	-10.969***	-11.370***
	(0.741)	(0.424)
Log(Assets)	0.885***	0.916***
	(0.053)	(0.034)
High Regime		-0.531*
		(0.291)
GDP	0.018	0.011
	(0.034)	(0.039)
Bank-Cap	4.428***	4.103***
	(0.933)	(0.705)
Mean-Salary	-0.751	-1.642*
	(0.913)	(0.841)
Log(δ)	2.097***	1.085***
	(0.634)	(0.417)
Zero-inflated model		

Intercept	1.176	-4.580*
	(1.681)	(2.712)
Log(Assets)	-0.176	-0.149
	(0.120)	(0.202)
High Regime		7.888***
		(2.502)
GDP	0.001	2.734***
	(0.109)	(0.787)
Mean-Salary	1.466	-48.468**
	(2.569)	(23.625)
AIC	1640.089	1597.558
Log Likelihood	-810.044	-786.779
Log-Likelihood ratio test		
- Statistic		46.530
- p.value		0.000
Number of observations	4329	4329

Note: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

5 Backtesting and robustness analysis

This section has a dual objective. First we want to construct a backtesting procedure for our model with regimes to determine its validity.[6] We also want to measure the extent that ignoring the existence of additional regimes in our operational loss data biases calculation of reserve capital if this reality is not formally considered.

5.1 Operational loss capital

The period selected to calculate coverage is January 2010 to December 2010. This period will be denoted by *C.* The regime is high for the month of January and low for the 11 other months. We number our two models as follows: Model 1 base model and Model 2 Markov regime + interaction with Business Lines and Event Types. To extend Dahen and Dionne (2010b), we construct our backtesting by taking into account regimes detected. Out-of-Sample backtesting

[6] We do not consider the model N+SN (Model 2) of Table 4 because it is inferior to N+ST4 and N+N.

does not include the period covered in the history, which lasts from January 2001 to December 2009 (denoted by *H*). For each model, the data from the periods *H* and *C* are scaled according to the estimated coefficients in Table 9. For a given bank, scaling is based on the mean value of *log (Assets)* of the bank during period *C*. Once scaled for a given bank, the historical losses (*H*) can be considered to follow a lognormal distribution. If we consider the bank U.S. Bancorp, the Kolmogorov-Smirnov test gives a statistic of D = 0.1328 and p-value = 0.1979. Because the lognormal law is the null hypothesis, the test does not allow us to reject it. Given the linearity in log (Assets) of the two models, we can conclude that the lognormal is valid for all banks in our BHC sample. We estimate the frequency according to Table 10.

We start with a random sample of 2,000 observations from the NB distribution. For each of the 2,000 numbers or counts, we draw random observations obtained from the lognormal distribution to compute the convolutions. By repeating the experience 100 times, we generate 200,000 observations for the backtest. This gives us a distribution for which we calculate the reserve capital (VaR) for four degrees of confidence: 95%, 99%, 99.5% and 99.9%. The 99.5% degree of confidence lets us evaluate the thickness of the distribution tail, and gives us an idea of what is happening when the VaR at 99.9% is not exceeded. Results are presented in tables 11 and 12.

Regarding statistical tests for the VaR, we performed the Kupiec (1995) test, which evaluates the number of values in excess of VaR, followed by the DQ test by Engle and Monganelli (2004) to measure the independence of these values. Lastly, the Christoffersen (1998) test helps us determine the conditional simultaneous coverage of frequency and independence of the values in excess of VaR.

We have 445 losses recorded for the period *H* and 63 for the period *C*, which gives us 508 losses. We must calculate the probable losses that a given bank incurs during period *C*. Accordingly, the 63 losses of *C* are scaled to the size of the bank, and each loss is multiplied 56 times by the scaling of the models to simulate all 8 *BusinessLines* and 7 possible *EventTypes* according to the Basel nomenclature (see Table 3). This lets us manage operational risk in all possible cases. The 63 losses therefore generate 3,528 possible losses, on which we perform statistical backtesting. The scaling covers all historical losses of *H*.

We perform the calculations for two banks. The first, in Table 11, is U.S. Bancorp (as in Dahen and Dionne, 2010b). Results show that the DQ test rejects the VaR at 99.5% for base Model 1 (no regime). The VaR is not rejected for the other tests at all degrees of confidence of Model 1. The bank's total assets are $290.6 billion, and reserve capital at 99.9% represents 1.02% of assets. Model 2 shows a small weakness in the frequency of values in excess of VaR at 99% but we do not reject the model at 1%. Regarding the estimated reserve of capital, it is lower in Model 2 than for the benchmark Model 1, $2,060.7 million for VaR at 99.9%.

We conclude with two important remarks. The first is that both models are validated by backtesting. The second is that capital calculated with Model 2 is below that calculated for Model 1. This affirms a temporal bias during the high-loss period. Using the calculation of Model 2, this bias for U.S. Bancorp is (2957.4-2060.7)/2957.4, which is very high, at 30.3%.

As further proof, we do the same process for a second BHC bank: Fifth Third Bancorp (Table 12). Its size is $111.5 billion. We obtain largely the same pattern. Model 2 is still the least capital expensive. If we consider Model 2 valid, the savings in reserve capital at 99.9% would be (1722.6-1291.5)/1722.6 = 25%. Further, the cross-loading of regimes with business lines and event types seems to capture the fact that these variables do not have the same effects during different phases of the regimes.

Table 11: Backtesting of U.S. Bancorp bank[1]

Backtesting	Confidence level	Frequency		VaR[2]	Kupiec test		DQ of E-M		Christoffersen	
		Theoretical	Observed		Stat.	p.value	Stat.	p.value	Stat.	p.value
Reference model	95%	0.050	0.043	269.7	1.760	0.1846	2.359	0.6701	2.550	0.2795
(Model 1)	99%	0.010	0.012	842.3	0.479	0.4887	1.404	0.8434	0.479	0.7869
	99.5%	0.005	0.008	1289.7	2.385	0.1225	14.792	**0.0052**	5.027	0.0810
	99.9%	0.001	0.002	2957.4	1.991	0.1583	2.751	0.6003	1.991	0.3696
HMM regimes and interactions										
(Model 2)	95%	0.050	0.043	209.0	1.760	0.1846	3.430	0.4886	1.844	0.3977
	99%	0.010	0.016	619.3	5.730	0.0167	9.258	0.0550	5.730	0.0570
	99.5%	0.005	0.004	913.6	0.116	0.7334	0.208	0.9949	0.116	0.9436
	99.9%	0.001	0.002	2060.7	0.666	0.4145	0.826	0.9349	0.666	0.7169

[1] Value in bold means model rejected at 1%.
[2] In millions of US dollars.

Table 12: Backtesting of Fifth Third Bancorp bank

Backtesting	Confidence level	Frequency		VaR[1]	Kupiec test		DQ of E-M		Christoffersen	
		Theoretical	Observed		Stat.	p.value	Stat.	p.value	Stat.	p.value
Reference model	95%	0.050	0.038	115.0	5.590	0.0181	7.547	0.1097	5.816	0.0546
(Model 1)	99%	0.010	0.007	430.7	1.553	0.2127	1.614	0.8063	1.553	0.4600
	99.5%	0.005	0.004	689.8	0.484	0.4867	0.511	0.9724	0.484	0.7851
	99.9%	0.001	0.003	1722.6	3.822	0.0506	5.812	0.2137	3.822	0.1479
HMM regimes and interactions										
(Model 2)	95%	0.050	0.042	94.4	2.783	0.0953	4.984	0.2889	2.804	0.2461
	99%	0.010	0.013	338.1	1.829	0.1762	3.369	0.4980	1.829	0.4007
	99.5%	0.005	0.007	522.6	1.570	0.2102	2.205	0.6981	1.570	0.4562
	99.9%	0.001	0.003	1291.5	6.057	0.0138	10.012	0.0402	6.057	0.0484

[1] In millions of US dollars.

5.2 Number of states in HMM model

To provide a robustness analysis of our research, we raise two questions. First, can we statistically justify that a combination of two normals, instead of one normal and an ST4, would have been insufficient? The second question is whether the regime should have three levels rather than two. A three-level regime would be a mixture of two normals plus an ST4 (skew-t type 4). To summarize, we compare our model N+ST4 with two other models: N+N and 2N+ST4. [7] We tested the normality of the pseudo-residuals of the three models as shown in Table 13. First, concerning the model N+N with two levels, all three p-values are below 10%. The data clearly show that this model is inadequate. Regarding the three-level model 2N+ST4, we have p-values of 0.0559, 0.0678 and 0.1863 for Kolomogorov-Smirnov, Anderson-Darling and Shapiro-Wilk respectively. If we apply the tests at 10%, two tests reject the three-level model, whereas only Anderson-Darling indicates a problem for the two-level N+ST4. In addition, the value of the AIC criterion of the model 2N+ST4 is 325.59 versus 319.68 (Table 4) for our two-level model N+ST4, which implies deterioration in performance for the 2N+ST4 model. This deterioration is more evident when we use the BIC criterion, which becomes 380.66 for the three-level model, compared with 350.34 for the model N+ST4. We therefore reject the three-level model 2N+ST4 at a level of confidence of 10%. Consequently, we retain the two-level specifications with a normal law in the low state and one skew t type 4 for our extreme observations.

Table 13: Statistical tests on pseudo-residuals presuming the existence of a 3-level model

		(1) 3-level model 2N+ST4		(2) 2-level model N+ST4		(3) 2-level model N+N	
1	Kolmogorov-Smirnov	0.0819	0.0559	0.0718	0.1540	0.0849	0.0408
2	Anderson-Darling	0.6950	0.0678	0.6940	0.0682	0.7873	0.0400
3	Shapiro-Wilk	0.9839	0.1863	0.9831	0.1650	0.9790	0.0678

[7] The estimation results of the model 2N+ST4 are not presented here but are available from the authors.

Conclusion

In this article, we demonstrated that considering business cycles can reduce capital (at 99.9%) for operational risk by redistributing it between high regime and low regime states. The variation of capital is estimated to be between 25% and 30% in our period of analysis. We also provide evidence that court settlements significantly affect the temporal distribution of losses. Several large losses were reported after the financial crisis of 2007-2009 owing to these delays. This phenomenon is not new; it has also been observed in the insurance industry.

Several extensions of our study are possible. The most promising would be to test the stability of the results using different regime detection methods (Maalaoui, Chun et al., 2014). Specifically, an effective real-time regime detection approach should consider the asymmetry observed in this article. This approach would notably allow separate analysis of level and volatility regimes.

Another possible extension is to use a different approach than the scaling of operational losses. Some banks use the Change of Measure Approach proposed by Dutta and Babbel (2013). It would be interesting to examine whether the results of this approach can remain stable by introducing cycles in the data.

References

Allen, L. and Bali, T.G. (2007). Cyclicality in catastrophic and operational risk measurements. Journal of Banking and Finance 31, 1191-1235.

Ames, M., Schuermann, T., and Scott, H.S. (2014). Bank capital for operational risk: A tale of fragility and instability. Available at SSRN 2396046.

Bhamra, H.S., Kuehn, L.A., and Strebulaev, I.A. (2010). The levered equity risk premium and credit spreads: A unified framework. Review of Financial Studies 23, 645-703.

Carrillo, C., Gzyl, H., and Tagliani, A. (2012). Reconstrructing heavy-tailed distributions by splicing with maximum entropy in the mean. Journal of Operational Risk 7(2), 1-13.

Chapelle, A., Crama, Y., Hübner, G., and Peters, J.P. (2008). Practical methods for measuring and managing operational risk in the financial sector: a clinical study. Journal of Banking and Finance 32, 1049-1061.

Chavez-Demoulin, V., Embrechts, P., and Nešlehová, J. (2006). Quantitative models for operational risk: extremes, dependence and aggregation. Journal of Banking and Finance 30, 2635-2658.

Chernobai, A. and Yildirim, Y. (2008). The dynamics of operational loss clustering. Journal of Banking and Finance 32(12), 2655-2666.

Christoffersen, P.F. (1998). Evaluating interval forecasts. International Economic Review 39(4), 841-862.

Chen, H. (2010). Macroeconomic conditions and the puzzles of credit spreads and capital structure. Journal of Finance 65, 2171-2212.

Cope, E. and Labbi, A. (2008). Operational loss scaling by exposure indicators: evidence from the ORX database. Journal of Operational Risk 3(4), 25-45.

Cummins, D.J., Dionne, G., McDonald, J.B., and Pritchett, M.B. (1990). Application of the GB2 family of distributions in modeling insurance loss processes. Insurance: Mathematics and Economics 9, 257–272.

Cruz, M.G. (2002). Modeling, Measuring and Hedging Operational Risk. John Wiley & Sons, Chichester.

Dahen, H. and Dionne, G. (2010b). Scaling models for the severity and frequency of external operational loss data. Journal of Banking and Finance 34(7), 1484-1496.

Dahen, H., Dionne, G. and Zajdenweber, D. (2010a). A practical application of extreme value theory to operational risk in banks. Journal of Operational Risk 5(2), 63-78.

Dionne, G. (1992). Contribution to Insurance Economics, Kluwer Academic Press, Boston.

Dionne, G. and Saissi Hassani, S. (2016). Endogenous hidden markov regimes in operational loss data: Application to the recent financial crisis. Working paper 15-03, Canada Research Chair in Risk Management. http://chairegestiondesrisques.hec.ca/en/wp-content/uploads/sites/2/2015/02/15-03.pdf.

Dutta, K.K. and Babbel, D.F. (2014). Scenario analysis in the measurement of operational risk capital: A change of measure approach. Journal of Risk and Insurance 81(2), 303-334.

Engle, R.F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics 22(4), 367-381.

Fernandez, C., Osiewalski, J., and Stell, M.F.J. (1995). Modeling and inference with v-spherical distributions. Journal of the American Statistical Association 90(432), 1331-1340.

Frachot, A. and Roncalli, T. (2002). Mixing internal and external data for managing operational risk. Working paper. Groupe de Recherche Opérationnelle, Crédit Lyonnais.

Hartung, T. (2004). Operational risks: modelling and quantifying the impact of insurance solutions. Working paper. Ludwig-Maximilians-University.

Jarrow, R.A. (2008). Operational risk. Journal of Banking and Finance 32(5), 870-879.

Kupiec, P.H. (1995). Techniques for verifying the accuracy of risk measurement models. The Journal of Derivatives 3(2), 73-84.

Lewis, M.A. (2003). Cause, consequences and control: towards a theoretical and practical model of operational risk. Journal of Operations Management 21, 205-224.

Maalaoui Chun, O., Dionne, G., and Francois, P. (2014). Detecting regime shifts in credit spreads, Journal of Financial and Quantitative Analysis 49(5-6), 1339-1364.

McDonald, J.B. and Xu, Y.J. (1995). A generalization of the beta distribution with applications. Journal of Econometrics 66, 133–152.

Na, H.S., van den Berg, J., Couto Miranda, L., and Leipoldt, M. (2006). An econometric model to scale operational losses. The Journal of Operational Risk 1(2), 11-31.

Nešlehová, J., Embrechts, P. and Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for operational risk. Journal of Operational Risk 1(1), 3-25.

Rigby, B., Stasinopoulos, M., Heller, G. and Voudouris, V. (2014). The distribution toolbox of GAMLSS. http://www.gamlss.org/wp-content/uploads/2014/10/distributions.pdf.

Shih, J., Samad-Khan, A., and Medapa, P. (2000). Is the size of an operational loss related to firm size? Operational Risk 2, 1-2.

Visser, I. and Speekenbrink, M. (2010). depmixS4: An R-package for hidden Markov models. Journal of Statistical Software 36(7), 1-21.

Wei, R. (2007). Quantification of operational losses using firm-specific information and external database. The Journal of Operational Risk 1(4), 3-34.

Zucchini, W. and MacDonald, I.L. (2009). Hidden Markov models for time series: An introduction using R. CRC Press, 276 p.

Appendix 1: The Baum-Welch algorithm

We start with a vector of initial arbitrary values $\theta^{(0)}$. EM is an iterative process. Each iteration is made up of two steps, E and M. For each iteration (k), step E is to calculate a function *Q* defined as the mathematical expectation of the log probability, if we know the sequence $x_{1:T}$ and using the value of the parameters $\theta^{(k)}$ such that:

$$Q\left(\theta,\theta^{(k)}\right)=E_{\theta^{(k)}}\left[log\,Pr\left(x_{1:T},s_{1:T}\middle|z_{1:T},\theta\right)\middle|x_{1:T},\theta^{(k)}\right]. \quad \text{(A.1)}$$

Then, in step M, we look for the value of the vector θ that maximizes $Q\left(\theta,\theta^{(k)}\right)$. This gives us a new set of parameters to find, namely:

$$\theta^{(k+1)}=arg\max_{\theta}Q\left(\theta,\theta^{(k)}\right). \quad \text{(A.2)}$$

$\theta^{(k+1)}$ will be compared with $\theta^{(k)}$ to verify the convergence criteria. In the absence of convergence, $\theta^{(k+1)}$ will serve as an entry for the following iteration $k+1$, and so on. The Baum-Welch algorithm has been shown to always converge (Rabiner, 1989).

Because it is a mathematical expectation, the quantity *Q* corresponds to computing a weighted sum of all of the possible probabilities for each of the three members to the right of equation (3.5) in the paper. This gives:

$$\begin{aligned} Q\left(\theta,\theta^{(k)}\right)=&\textstyle\sum_{j=1}^{n}\gamma_1(j)\,log\,Pr\left(s_1=j\middle|z_1,\theta_0\right)\\ &+\textstyle\sum_{t=2}^{T}\sum_{j=1}^{n}\sum_{k=1}^{n}\delta_t(j,k)\,log\,Pr\left(s_t=k\middle|s_{t-1}=j,z_{t-1},\theta_1\right)+\\ &\textstyle\sum_{t=1}^{T}\sum_{j=1}^{n}\gamma_t(j)\,log\,Pr\left(x_t\middle|s_t=j,z_t,\theta_2\right) \end{aligned} \quad \text{(A.3)}$$

where functions δ_t and γ_t represent the weights to calculate the mathematical expectation. Using the notation $M=\left\{z_{1:T},\theta^{(k)}\right\}$ to simplify the expressions, these weights δ_t and γ_t are written as:

$$\delta_t(j,k)=Pr\left(s_{t+1}=k,s_t=j\middle|x_{1:T},M\right) \quad \text{(A.4)}$$

$$\gamma_t(j)=Pr\left(s_t=j\middle|x_{1:T},M\right) \quad \text{(A.5)}$$

To calculate the probabilities δ_t and γ_t, let us define two probabilities α_t and β_t such that for all $i=1$ to *n* regimes:

$$\alpha_t(i)=Pr\left(x_{1:t},s_t=i|M\right) \tag{A.6}$$

$$\beta_t(i)=Pr\left(x_{t+1:T}|s_t=i,M\right). \tag{A.7}$$

In the literature, α_t is called a forward probability because of the relationship of recurrence $\alpha_t(j)=\left[\sum_i \alpha_{t-1}(i)a_{ij}\right]f_j(x_t)$. Similarly, β_t is called a backward probability because of the relationship:

$$\beta_t(i)=\left[\sum_j \beta_{t+1}(j)a_{ij}\right]f_j(x_{t+1}) \text{ with } \beta_T(i)=1\,\forall i.$$

The derivation of these relationships with vector notation is almost immediate, as in Zucchini and MacDonald (2009), by writing the probability function:

$$L_T=Pr\left(x_{1:T}|M\right)=\pi B_1A_2B_2...A_tB_t...A_TB_T1'. \tag{A.8}$$

By cutting the cross-product of equation (A.8) at time *t*, we have $\alpha_t=\pi B_1A_2B_2...A_tB_t$ and $\beta_t=A_{t+1}B_{t+1}...A_TB_T1'$ (with $\beta_T^{'}=1'$). Hence $\alpha_t=\alpha_{t-1}\times A_tB_t$ and $\beta_t=A_{t+1}B_{t+1}\times\beta_{t+1}$, which is the equivalent, in matrix notation, of the preceding forward and backward recurrence relationships. Now that our vectors α_t and β_t have been calculated, we can calculate the weight δ_t given that $\delta_t(j,k)=\alpha_t(j)\times f_k(x_{t+1})\times\beta_{t+1}(k)\times a_{jk}/\alpha_T1'$ as derived here:

$$\delta_t(j,k) \quad =Pr\left(s_{t+1}=k,s_t=j|x_{1:T},M\right)$$

$$=Pr\left(s_{t+1}=k,s_t=j,x_{1:T}|M\right)/Pr\left(x_{1:T}|M\right) \tag{A.9}$$

$$=Pr\left(x_{1:t},x_{t+1},x_{t+2:T},s_{t+1}=k,s_t=j,x_{1:T}|M\right)/L_T \tag{A.10}$$

$$=Pr\left(x_{1:t},s_t=j|M\right)Pr\left(x_{t+1},x_{t+2:T},s_{t+1}=k|x_{1:t},s_t=j,M\right)/L_T \tag{A.11}$$

$$=Pr\left(x_{1:t},s_t=j|M\right) \tag{A.12}$$

$$\times Pr\left(x_{t+1}|x_{t+2:T},s_{t+1}=k,x_{1:t},s_t=j,M\right) \tag{A.13}$$

$$\times Pr\left(x_{t+2:T}|s_{t+1}=k,x_{1:t},s_t=j,M\right) \tag{A.14}$$

$$\times Pr\left(s_{t+1}=k \middle| x_{1:t}, s_t=j, M\right) \Big/ L_T \tag{A.15}$$

$$=\alpha_t(j)\times f_k\left(x_{t+1}\right)\times \beta_{t+1}(k)\times a_{jk} \Big/ \alpha_T 1' \tag{A.16}$$

Equation (A.9) is obtained by simple application of Bayes' theorem. In (A.10) the sequence $x_{1:T}$ is cut into three pieces: from $x_{1:t}, x_{t+1:t+1}$ and $x_{t+2:T}$ using $L_T = Pr\left(x_{1:T} \middle| M\right)$ defined in (A.8). Equation (A.11) and equations (A.12) to (A.15) also use the Bayes model. Equation (A.12) is the direct expression of $\alpha_t(j)$. Equation (A.13) is simplified to $Pr\left(x_{t+1} \middle| s_{t+1}\right)$ because $x_{t+1} \middle| s_{t+1}$ is known independently from $x_{t+2:T}$ and from s_t (by the very construction of the HMM). In equation (A.14), the sequence $x_{t+2:T} \middle| s_{t+1}$ is independent from $x_{1:t}$ and from s_t. Lastly, on line (A.15), because $s_{t+1} \middle| s_t$ do not depend on $x_{1:t}$, the expression is reduced to $Pr\left(s_{t+1}=k \middle| s_t=j, M\right)$ which is equal to a_{jk} in (A.16). It now remains to be shown that $L_T=\sum_j \alpha_T(j)=\alpha_T 1'$. Based on definition (A.6) applied to $t=T, \alpha_T(i)=Pr\left(x_{1:T}, s_T=i \middle| M\right)$, the sum of $\alpha_T(i)$ on all *i* possible states must give the probability $Pr\left(x_{1:T} \middle| M\right)$, because the system is necessarily and exclusively in one or the other of the *i* states. The same reasoning permits us to find $\gamma_t(j)$ in function of δ_t noticing that

$$Pr\left(s_t=j \middle| x_{1:T}, M\right)=\sum_k Pr\left(s_{t+1}=k, s_t=j \middle| x_{1:T}, M\right).$$

Hence,

$$\gamma_t(j)=\sum_k \delta_t(j,k). \tag{A.17}$$

To summarize the construction of probabilities α_t and β_t , we first calculate δ_t which in turn yields γ_t. From this point, we can calculate the function $Q\left(\theta, \theta^{(k)}\right)$ to find $\theta^{(k)}=\theta$ which maximizes *Q*. This advances the EM process until convergence to obtain the vector θ of the final application parameters of the HMM. For our estimation, we have used the functions available in the package depmixS4 (Visser and Speekenbrink, 2010) with the skew t type 4 function of the gamlss package (Rigby et al, 2014), in R language by r-project.org.

Appendix 2: Tables and figures for HMM: N+EGB2

We also considered the GB2 distribution for the high regime while continuing with the Normal distribution for the low regime. The GB2 distribution is very general, having as particular cases many well-known distributions such as the Generalized Gamma, log t-student, Burr-3, Burr-12, Weibull and many others (Cummins et al., 1990; McDonald and Xu, 1995). This is a four-parameter distribution, well accepted for estimating the asymmetry of a distribution and the fatness of the tail.

Because our data are in logarithms, we must consider the exponential GB2 (EGB2). For $y \in (-\infty, +\infty)$, the density of the EGB2 is equal to:

$$f(y|a,b,p,q) = \frac{e^{pz}}{|b| \times Beta(p,q) \times (1+e^{z})^{p+q}}$$

where $z = \frac{y-a}{b}$, a and $b \in \mathbb{R}$, p and q > 0, $|b|$ is the absolute value of b and Beta is the Beta function. The EGB2 is positive asymmetric when p > q and negative asymmetric when p < q. The estimation results of the EGB2 parameters are presented in Table A1. According to the log Likelihood and AIC criteria, the EGB2 is equivalent to the N + ST4 distribution.

Table A1: Parameters estimations of HMM with N +EGB2

	Model 3 $N + ST4$		Model 4 $N + EGB2$	
	Variable	Coefficient	Variable	Coefficient
	Intercept	0.9772 (0.8161)	Intercept	0.7468 (0.7982)
	L_1	-1.7371*** (0.2798)	L_1	-1.6572*** (0.3088)
	Intercept	-25.7285*** (4.5707)	Intercept	-26.2582*** (4.9796)
	L_2	11.7434*** (2.4739)	L_2	12.0088*** (2.5634)
Low regime	Normal law		Normal law	
	μ_1	2.4172*** (0.5876)	μ_1	2.4197*** (0.6840)
	σ_1	0.7653*** (0.2006)	σ_1	0.7682*** (0.2075)
High regime	ST4		EGB2	
	μ_2	3.7872*** (0.5449)	a	0.8725** (0.4161)
	$\log(\sigma_2)$	-0.0415 (0.2546)	b	1.0359*** (0.2494)
	$\log(\nu)$	2.7734* (1.4299)	$\log(p)$	2.6096*** (0.4235)
	$\log(\tau)$	0.9492 (0.8007)	$\log(q)$	0.0927 (0.2427)
Log likelihood		-148.838	-148.785	
AIC criteria		319.677	319.570	
Number of observations		120	120	

Note: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Figure A1 shows that the tail of the ST4 remains higher than those of the other distributions to the right of the log (loss) equal to 8 on the x axis.

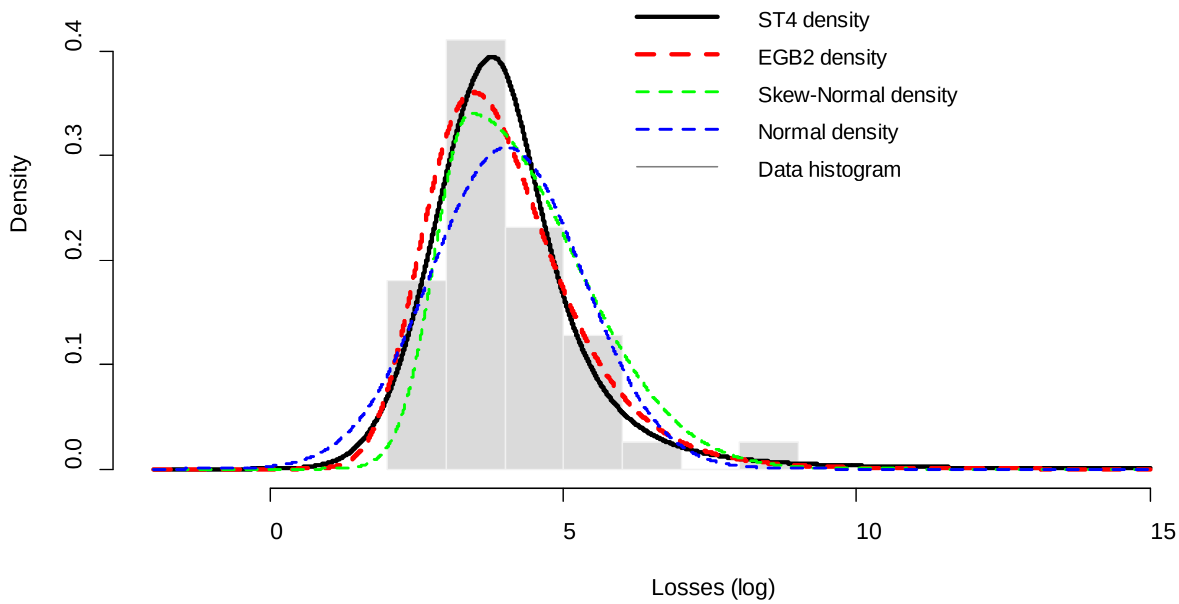

Figure A1: Data histogram and density of log of losses from different distributions

Another criterion used to retain the ST4 is related to the normality of the pseudo-residuals, as shown in Table 8 of the paper. Table A2 presents the results of the three tests for the EGB2.

Table A2: Normality tests of pseudo-residuals of HMM using N+EGB2

	Test	Pseudo-residuals	
		Statistic	p-value
1	Kolmogorov-Smirnov	0.0818	0.1206
2	Anderson-Darling	0.7937	0.0391
3	Shapiro-Wilk	0.9904	0.1133

Here again, the results are very similar to the ST4, although the Anderson-Darling test rejects normality even if the p-value is not far from 5%. The ST4 was not rejected at 5% with the three tests. Finally, we have estimated the skewness and the kurtosis obtained from different distributions. Results in Table A3 show that the ST4 provides the best results.

Table A3: Skewness and kurtosis analysis with different distributions in the high regime

	Skewness	Kurtosis
Data (log)	1.192	5.974
N	0.000	3.000
SN	0.653	3.319
ST4	1.119	6.127
EGB2	1.008	5.093