Centre Interuniversitaire sur le Risque,
les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **15-16**

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

Georges Dionne
Samir Saissi Hassani

Août/August 2015

Dionne: Canada Research Chair in Risk, HEC Montréal, CIRRELT and CIRPÉE
georges.dionne@hec.ca
Saissi Hassani: Canada Research Chair in Risk, HEC Montréal, CIRRELT and CIRPÉE
samir_saissi@yahoo.com

The authors thank FCI and SSHRC Canada for financial support.

Abstract:
We determine whether there is an endogenous Hidden Markov Regime (HMR) in the operational loss data of banks from 2001 to 2010. A high level regime is marked by very high loss values during the recent financial crisis. There is therefore temporal heterogeneity in the data. If this heterogeneity is not considered in 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 periods of a high regime. Variation in capital can go up to 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, bank's regulatory capital, Basel regulation

JEL Classification: G21, G24, G28

1 Introduction

Since the inception of operational risk modeling, authors have regularly highlighted the fact that the amount of reserve capital calculated is very fragile, even unstable. Ames, Schuermann and Scott (2014) clearly show this fragility with operational loss data related to the recent financial crisis that began in 2007.

Before that, 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, 2010). These studies argue for more conventional base models to better estimate the distributions and consider the presence of switching regimes in the data endogenously. In this paper, we build on the scaling model of Dahen and Dionne (2010) by detecting and incorporating endogenous Hidden Markov regimes for losses of one million dollars and more.

We show that the operational loss data of American banks are indeed characterized by a Hidden Markov 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 verify that their presence significantly affects the distribution of losses in general. These results are particularly important for some operational losses, particularly those linked to financial product pricing errors, over which several large banks have been sued during and after the recent financial crisis. We also analyse the scaling of the data to banks of different sizes and risk exposures, and present the results of backtesting of the model in different banks.

The general message of our contribution is that there is 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 a high regime period. Overall banks used too much capital for operational risk when the regimes are not considered in our period of analysis.

In Section 2, we present the database used. Section 3 discusses identification models of regimes and presents their estimation. 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. 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 three Tables: 1, 2, and 4.

Table 1 presents the size distribution of banks with $1 billion or more in assets that sustained 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 for the Basel regulation. Table 4 is a cross-loading table linking Event Types and Business Lines. We note that the largest mean losses are in Corporate Finance, Retail Brokerage and Trading and Sales for Business Lines, and in Clients, Products and Business Practice, Damage to Physical Assets, and Execution Delivery and Process Management for Event Types.

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 for standard deviation.

Table 2: Operational losses of BHC banks with bank asset 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 for standard deviation.

Table 3: Nomenclature of Event Types and Business Lines codes

Variables	Codes
Event Types	
Clients products and business practice	CliPBP
Business disruption and system failure	BusDSF
Damage to physical asset	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

Table 4: Cross-loading table of types of losses and business lines

Business lines		CliBP	BusDSF	DamPA	EmpPWS	EF	IF	ExeDPM	All
Rbr	Mean	28.6			18.7	78.8	8.0	2.8	22.8
	Sd	89.7			33.8		8.3	1.9	71.1
	Sum	1,030.3	0.0	0.0	149.7	78.8	103.5	5.7	1,367.9
	Count	36	0	0	8	1	13	2	60
PayS	Mean	62.4	19.2	743.0		23.9	18.7	11.1	67.5
	Sd	85.7	24.2			4.6	16.6	8.8	150.0
	Sum	873.7	57.6	743.0	0.0	47.8	56.1	44.6	1,822.7
	Count	14	3	1	0	2	3	4	27
TraS	Mean	91.6		55.0	6.9	18.0	130.7	139.1	95.7
	Sd	195.8			8.5		228.9	286.6	202.6
	Sum	3,756.1	0.0	55.0	34.7	18.0	1,437.5	1,113.0	6,414.4
	Count	41	0	1	5	1	11	8	67
ComB	Mean	45.4		1.0	16.0	18.6	22.1	36.7	26.8
	Sd	65.5			18.1	28.5	26.8	29.0	42.2
	Sum	1,045.3	0.0	1.0	80.1	725.7	309.4	147.0	2,308.5
	Count	23	0	1	5	39	14	4	86
RBn	Mean	369.8	2.0	1.0	7.3	17.8	11.0	29.4	147.2
	Sd	1,531.0			9.9	49.9	20.9	30.9	946.8
	Sum	21,819.8	2.0	1.0	58.8	604.6	505.4	264.3	23,255.9
	Count	59	1	1	8	34	46	9	158
AgnS	Mean	85.2				5.5	3.6		58.5
	Sd	138.2				6.5			117.2
	Sum	681.7	0.0	0.0	0.0	16.5	3.6	0.0	701.9
	Count	8	0	0	0	3	1	0	12
CorF	Mean	556.5					56.2	9.6	441.4
	Sd	1,473.2					77.0	13.4	1,311.7

	Sum	21,148.3	0.0	0.0	0.0	0.0	449.5	28.7	21,626.5
	Count	38	0	0	0	0	8	3	49
AssM	Mean	75.3				95.0	61.6	37.4	72.2
	Sd	153.3				128.7	94.1	50.0	141.8
	Sum	3,012.2	0.0	0.0	0.0	189.9	184.9	149.4	3,536.5
	Count	40	0	0	0	2	3	4	49
All	Mean	206.1	14.9	200.0	12.4	20.5	30.8	51.6	120.1
	Sd	941.9	21.5	362.9	21.0	42.3	88.0	143.2	680.7
	Sum	53,367.3	59.6	800.0	323.3	1,681.4	3,049.9	1,752.8	61,034.3
	Count	259	4	4	26	82	99	34	508

Note: Loss amounts are in millions of dollars.

3 Identification of regimes

We assume that there are regimes in operational loss data. To support this assertion, we present Figures 1, 2, and 3. 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. We also note that 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 for having marketed complex financial products that were poorly structured, with incorrect prices and dubious ratings. Figure 2 presents similar evolutions in loss volatility. Figure 3 shows that the trend for number of losses of one million dollars and more is a sawtooth, but there is no major increase during and after the financial crisis. The year 2003 exhibits the highest frequencies.

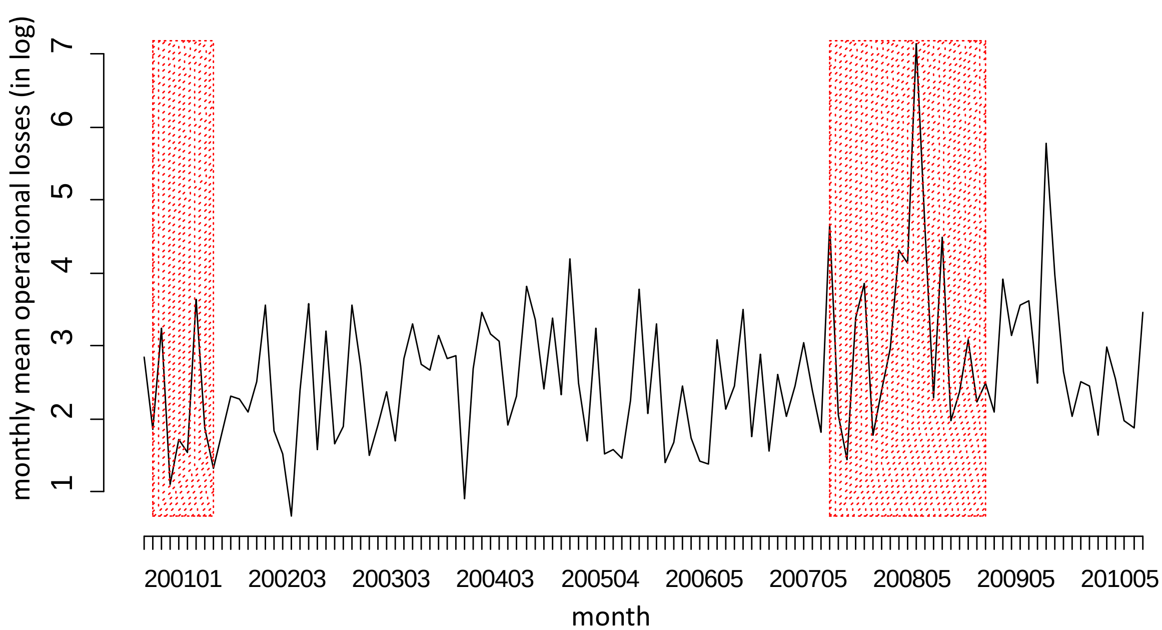

Figure 1: Changes in monthly mean operational losses

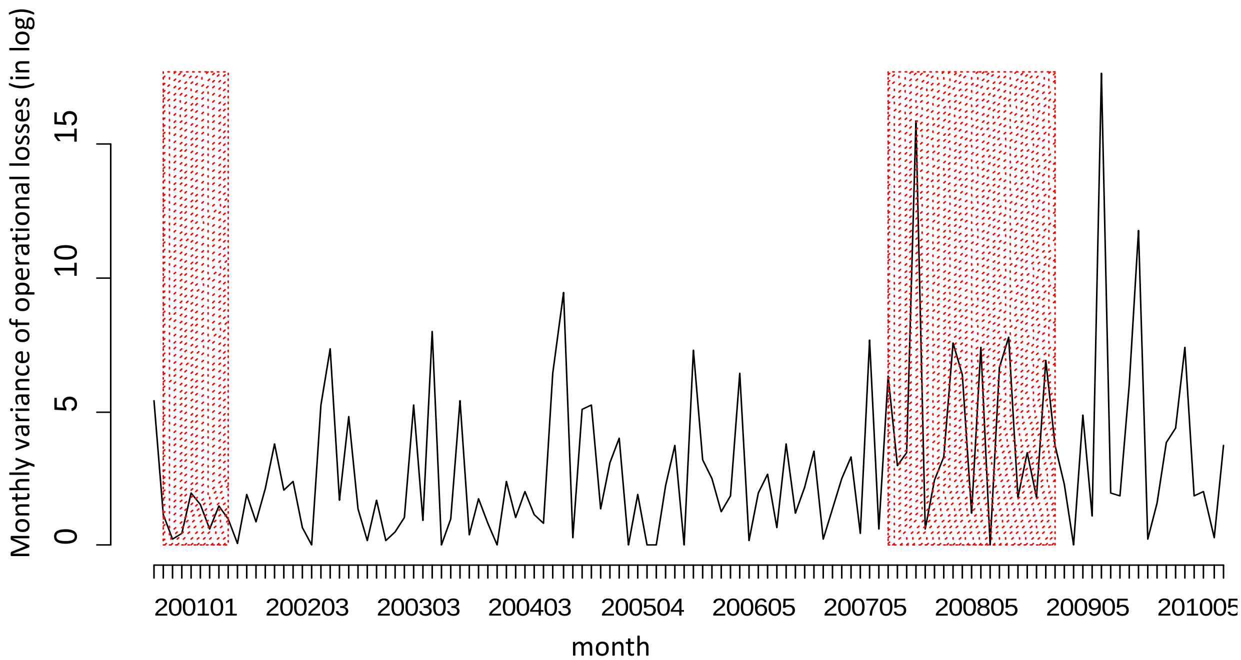

Figure 2: Changes in monthly variance of operational losses

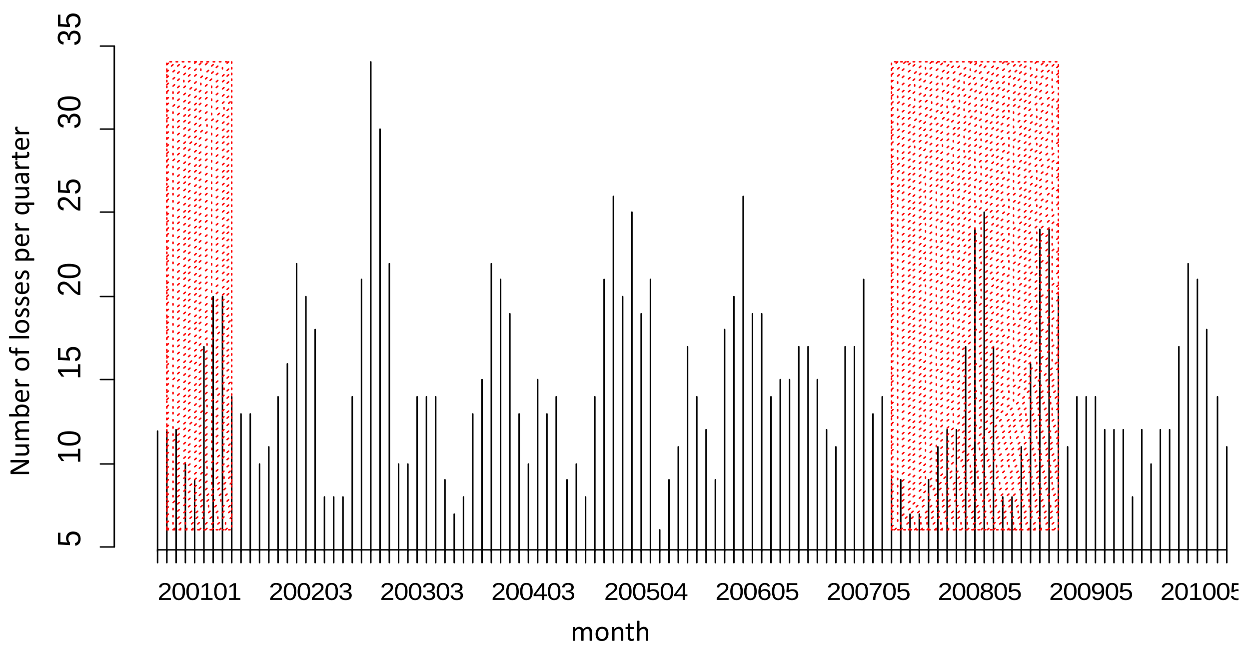

Figure 3: Changes in number of operational losses

3.1 Markov Switching Regimes

3.1.1 Literature

Several researchers have attempted to detect the presence of unobservable regimes by using a Markov process (Hamilton, 1989; Rabiner, 1989). Since then, increasingly rich developments of the model have emerged in all fields of research. Siu (2007) shows the advantage of applying this methodology in finance and actuarial science to better price insurance products. Korolkiewicz and Elliott (2007) propose a credit rating model based on the concept of Markov Switching. Siu and Yang (2007) model the Conditional Value at Risk (CVaR) advantageously for market and credit risk models using a complete procedure. Liechty (2013) presents another example of Markov Switching as a risk management tool. The origins of HMM (Hidden Markov Modeling) date back to the 1960s, with Baum and Petrie (1966) and Baum et al. (1970). Hamilton (1989) made a dual contribution: he paved the way for the use of HMM in economics and finance, and developed his own estimation method called the Hamilton Filter. This method is very useful in cases where different levels of the regime are modeled with normal distributions.

The Hamilton Filter implicitly supposes that observations come from distributions with a sufficient number of draws to notably consider that the initial conditions describing the system at starting time t = 1 has only a small effect on its evolution. This hypothesis has been studied in depth by Psaradakis and Sola (1998), who show that one would need a sample of at least 400 observations to guarantee that the estimate works well, especially in the presence of known fat-tailed data. For this reason, we use the Baum-Welch algorithm, which we describe below, to estimate our model. As Mitra and Date (2010) and Bulla (2011) showed, this algorithm does not use a priori assumptions of distributions.

3.1.2 Markov Switching Model

The basic idea behind this model is intuitive. We suppose that the data 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. There are two probabilities that describe each state. 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 endogenously (state 1). We inject information of loss severity and frequency that comes uniquely from the data, such that the model will show the unobservable underlying dynamics. We also analyse a three-state application in the robustness section of the paper.

3.1.3 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 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 $t, s_t \in \{1,...,n\}$. We 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 covariables noted as z_t. According to Proposition 2 of Mitra and Date (2010), a HMM is well defined when the parameters $\{A,B,\pi\}$ are known, *A* being the transition matrix $n \times n$ whose elements are written as $a_{ij} = Pr\left(s_{t+1} = j \middle| s_t = i, z_t, \theta\right)$, *B* is a diagonal matrix whose elements $b_i\left(x_t\right) = Pr\left(x_t \middle| s_t, z_t, \theta\right)$ 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 $t=1$, $\pi_i = Pr\left(s_1 = i \middle| z_t, \theta\right)$, $\pi = \left(\pi_1,...,\pi_i,...,\pi_n\right)$. 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). The choice of this mixture of distributions will be justified at the end of this section. For now, note that $B_t = \begin{bmatrix} f_1(x_t) & 0 \\ 0 & f_2(x_t) \end{bmatrix}$ such that:

$$f_1(x_t | \mu_1, \sigma_2) = \frac{1}{\sigma_1 \sqrt{2\pi}} exp\left[\frac{(x_t - \mu_1)^2}{2\sigma_1^2} \right] \tag{3.1}$$

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

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

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

where

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

B is the beta function $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ where Γ is the gamma function.

Concerning the matrix $A_t = \begin{bmatrix} (1-a_{12}) & a_{12} \\ (1-a_{22}) & a_{22} \end{bmatrix}$, the elements a_{ij} will be modeled according to the m independent covariables $z_t = (z_t^1, ..., z_t^m)$. We posit that:

$$a_{ij} = logistic(\eta_{ij} z_t) \tag{3.3}$$

where $logistic(\bullet)$ is the logistic function $\left(\frac{exp(\bullet)}{1+exp(\bullet)} \right), \eta_{ij} = (\eta_{ij,0}, ..., \eta_{ij,k} ..., \eta_{ij,m}), \eta_{ij,0}$ is a constant and $\eta_{ij,k}$ is the coefficient to estimate for the k^{th} covariable z_t^k relative to the conditional probability a_{ij}. Regarding the initial distribution π, a priori, it may depend on $z_{t=1} = z_1$. However, below we will estimate π as a vector of constants. We can separate the θ parameters 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 the 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 $\left(\theta_0 = (\pi_1, ..., \pi_n)\right)$.

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 in the HMM context. We start with a vector of initial arbitrary values $\theta^{(0)}$. EM is an iterative process. Each loop is made up of two steps, E and M. For each loop (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:

$$Q\left(\theta, \theta^{(k)}\right) = E_{\theta^{(k)}}\left[log\,Pr\left(x_{1:T}, s_{1:T} \middle| z_{1:T}, \theta\right) \middle| x_{1:T}, \theta^{(k)}\right]. \quad (3.6)$$

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:

$$\theta^{(k+1)} = arg\,\max_{\theta} Q\left(\theta, \theta^{(k)}\right). \quad (3.7)$$

$\theta^{(k+1)}$ 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 loop $k+1$, 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). This gives:

$$Q\left(\theta,\theta^{(k)}\right)=\sum_{j=1}^{n}\gamma_1(j)\,log\,Pr\left(s_1=j\,|\,z_1,\theta_0\right)$$
$$+\sum_{t=2}^{T}\sum_{j=1}^{n}\sum_{k=1}^{n}\delta_t(j,k)\,log\,Pr\left(s_t=k\,|\,s_{t-1}=j,z_{t-1},\theta_1\right)+$$
$$\sum_{t=1}^{T}\sum_{j=1}^{n}\gamma_t(j)\,log\,Pr\left(x_t\,|\,s_t=j,z_t,\theta_2\right)$$

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:

$$\delta_t(j,k)=Pr\left(s_{t+1}=k,s_t=j\,|\,x_{1:T},M\right) \tag{3.8}$$

$$\gamma_t(j)=Pr\left(x_{t+1:T}\,|\,s_t=i,M\right) \tag{3.9}$$

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):

$$\alpha_t(i)=Pr\left(x_{1:t},s_t=i\,|\,M\right) \tag{3.10}$$

$$\beta_t(i)=Pr\left(x_{t+1:t},s_t=i,M\right) \tag{3.11}$$

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

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

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

$$L_T=Pr\left(x_{1:T}\,|\,M\right)=\pi B_1A_2B_2...A_tB_t...A_TB_T1'\,. \tag{3.12}$$

By cutting the cross-product of equation 3.12 at time *t*, we have $\alpha_t = \pi B_1 A_2 B_2 ... A_t B_t$ and $\beta_t = A_{t+1} B_{t+1} ... A_T B_T 1'$ (with $\beta_T^{'} = 1'$). Hence $\alpha_t = \alpha_{t-1} \times A_t B_t$ 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 $\delta_t(j,k) = \alpha_t(j) \times f_k(x_{t+1}) \times \beta_{t+1}(k) \times a_{jk} / \alpha_T 1'$ as derived here:

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

$$= Pr\left(s_{t+1} = k, s_t = j, x_{1:T} | M\right) / Pr\left(x_{1:T} | M\right) \quad (3.13)$$

$$= Pr\left(x_{1:t}, x_{t+1}, x_{t+2:T}, s_{t+1} = k, s_t = j, x_{1:T} | M\right) / L_T \quad (3.14)$$

$$= Pr\left(x_{1:t}, s_t = j | M\right) Pr\left(x_{t+1}, x_{t+2:T}, s_{t+1} = k \middle| x_{1:t}, s_t = j, M\right) / L_T \quad (3.15)$$

$$= Pr\left(x_{1:t}, s_t = j | M\right) \quad (3.16)$$

$$\times Pr\left(x_{t+1} \middle| x_{t+2:T}, s_{t+1} = k, x_{1:t}, s_t = j, M\right) \quad (3.17)$$

$$\times Pr\left(x_{t+2:T} \middle| s_{t+1} = k, x_{1:t}, s_t = j, M\right) \quad (3.18)$$

$$\times Pr\left(s_{t+1} = k \middle| x_{1:t}, s_t = j, M\right) / L_T \quad (3.19)$$

$$= \alpha_t(j) \times f_k(x_{t+1}) \times \beta_{t+1}(k) \times a_{jk} / \alpha_T 1' \quad (3.20)$$

Equation (3.13) is obtained by simple application of Bayes' theorem. In (3.14) 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 $L_T = Pr\left(x_{1:T} | M\right)$ defined in (3.12). Equation (3.15) and equations (3.16) to (3.19) also use the Bayes model. Equation 3.16 is the direct expression of $\alpha_t(j)$. Equation (3.17) is simplified to $Pr\left(x_{t+1} \middle| s_{t+1}\right)$ because $x_{t+1} | s_{t+1}$ is known independently from $x_{t+2:T}$ and from s_t (by the very construction of the HMM). In equation (3.18), the sequence $x_{t+2:T} | s_{t+1}$ is independent from $x_{1:t}$ and from s_t. Lastly, on line (3.19), because $s_{t+1} | s_t$ 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 (3.20). It now remains to be shown that $L_T = \sum_j \alpha_T(j) = \alpha_T 1'$. Based on definition

(3.10) applied to $t=T, \alpha_T(i)=Pr(x_{1:T}, s_T=i|M)$, the sum of $\alpha_T(i)$ on all *i* possible states must give the probability $Pr(x_{1:T}|M)$, 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

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

Hence,

$$\gamma_t(j)=\sum_k \delta_t(j,k). \tag{3.21}$$

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 $Q(\theta,\theta^{(k)})$ 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.

Concretely, we construct the sequence $x_{1:T}$ from monthly mean losses (in log). We already know that the means are far from following a normal distribution. We consequently use a mixture where the first "normal" state will be modeled by a normal distribution and the second state of the high regime (abbreviated as 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 *lc*123 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. Because the transition matrix is not constant, our model can be called non homogeneous. In short, we use four distributions as follows:

$$
\begin{aligned}
x_t \left| s_t \right. &= 1 \sim N\left(\mu_1, \sigma_1\right) \\
x_t \left| s_t \right. &= 2 \sim ST4\left(\mu_2, \sigma_2, \nu, \tau\right) \\
a_{12} &= logistic\left(\eta_{12,0} + \eta_{12,1} lc123\right) \\
a_{22} &= logistic\left(\eta_{22,0} + \eta_{22,1} lc123\right).
\end{aligned}
\tag{3.22}
$$

Lastly:

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

3.1.4 Results and discussion

The results of the estimation of the model are presented in Table 5. We begin with the parameters of the two distributions that we use. 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 coefficients are very significant at all degrees of confidence chosen. Regarding the Skew-t type 4, its mean is estimated at 3.7872, whereas its standard deviation can be considered equal to 1 (its log can be considered statistically null because it is non-significant). In a high regime, we therefore have a significant and simultaneous increase in the mean and an increase in the standard deviation. In addition, the asymmetry of the Skew-t type 4 is confirmed by the *log*(*Shape.nu*) coefficient significant at 10%. We will return to the validation of these distributions below by performing a robustness analysis of our statistical results.

Table 5: Estimation of the Hidden Markov Model

Variable		Coefficient
Probability of transition to High Regime	Intercept	0.9772
	lc123	-1.7371***
Probability of staying in High Regime	Intercept	-25.7285***
	lc123	11.7434***
Estimation of Normal distribution	mu	2.4172***
	sigma	0.7653***
Estimation of ST4 distribution	mu	3.7872***
	log (sigma)	-0.0415
	log(shape.nu)	2.7734*
	log(shape.tau)	0.9492
Estimation of HMM model	Log max likelihood	-148.838
	AIC criteria	319.677
	Number of observations	120

Note: *indicates significant at 10% and *** indicates significant at 1%.

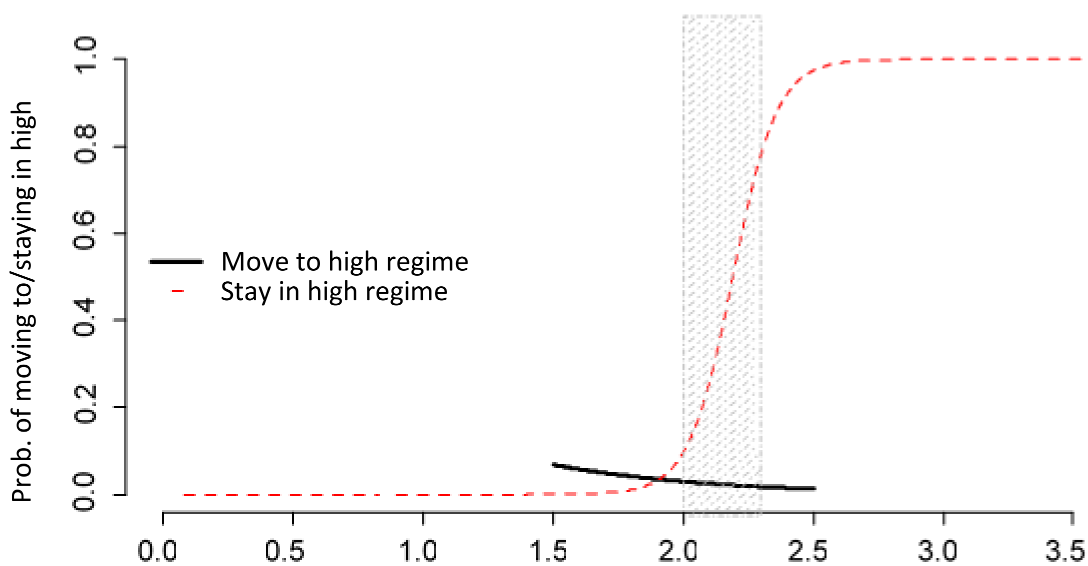

Figure 4: Markov transition probabilities

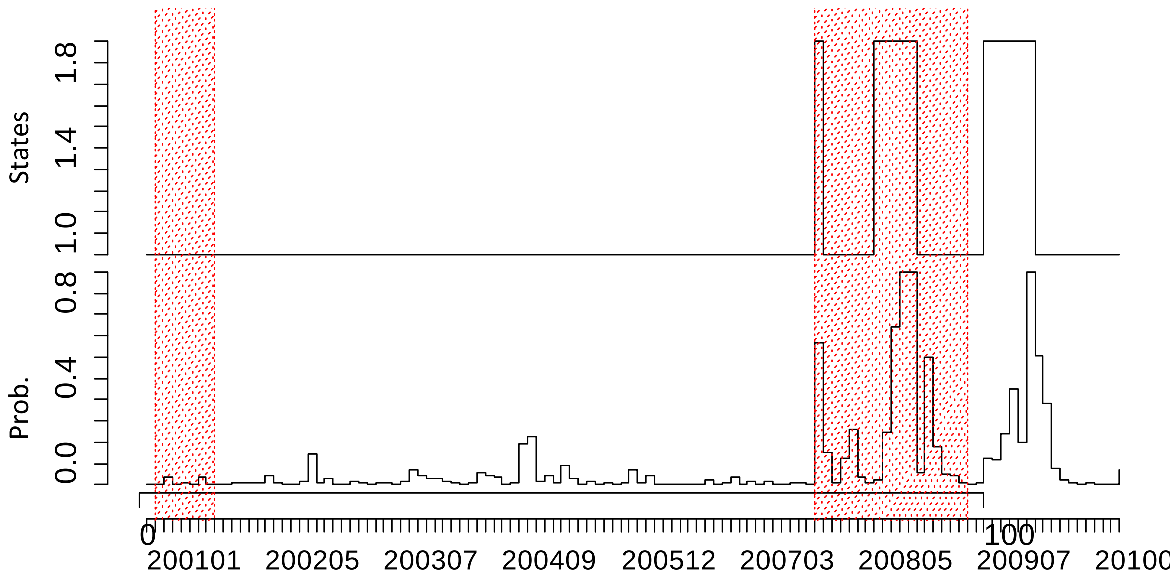

Figure 5: Markov Regimes detected from January 2001 to December 2010

The estimation of Table 5 gives a value of $\nu = exp(0.7734) = 2.167$ 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 5 is not significant, $log(\tau)$ can be considered null, therefore $\tau = 1$. The right distribution tail would be thicker in this sense. Given these estimated two degrees of freedom markedly below 30, this is confirmation that we are far from a normal law where $\nu > 30$ and $\tau > 30$.

We now discuss the stages of the transition probability in Table 5. The coefficient of the variable *lc*123 is very significantly negative. This means that the larger the number of losses, the lower the probability of starting from a high regime, which would be a bit odd. To understand what is happening, we draw in Figure 4 the curves of the two transition functions: move to or stay in a high regime. Note that the number of losses is historically limited to between 7 and 20 per quarter (where *lc*123 is included between 2 and 3). In this case, in Figure 4, the section to the left of the point *lc*123 = 2 would be meaningless, and was therefore cut from the figure. The part to the right of this point presents a barely declining curve, nearly parallel with the X axis, with a value of about 5% as a probability of moving to a high regime (HR). We can reasonably assume that the number of loss announcements does not play a role in predicting movement to a HR, nor does the increase (or not) in operational losses. Consequently, by reformulating the foregoing in statistical terms, we have found evidence to support the hypothesis of independence of distributions of frequencies and severities, which is an important contribution of this research. To continue with the probability of remaining in a high regime, if the number of loss announcements is between 7 and 12 per quarter, the mean probability of staying is about 50%. At between 12 and 24 mean quarterly losses, the probability of remaining in a high regime state is practically 99%.

Let us now consider Figure 5, which shows the Markov switching states detected. Three facts emerge from the figure. First, there was almost no reaction for 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 latter happens after the end of the recession.

It is interesting to document this fact by analyzing what happened for the two variations. To do so, we take the individual losses at the largest amounts, which represent at least 80% of the total lost during each period analyzed. We obtained information on what happened for 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 first loss was incurred by Wachovia Bank in July 2008. It comprises a series of final writedowns linked to mortgages. The class action suit filed in federal court in California on June 6, 2008 alleges that the bank distorted its standards for underwriting option adjustable rate mortgages (ARMs), with payment structures that lacked the usual guarantees that were nonetheless stipulated in the contracts. This is a CliPBP type loss. The second loss, for the same amount, i.e. $8.4 billion, concerns CFC of Bank of America. In October 2008, it was accused of illegal practices concerning products related to bank loans; 400,000 buyers were affected. CFC had to agree to settle the lawsuits filed against it by a group of attorneys general in 11 states, including California, Florida, Illinois, Connecticut, and Washington. 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. We will see that 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. We begin with Citigroup, which announced a loss of $840 million in January 2010. This loss results from an accounting error related to the way the bank calculated its CVA (Credit Value Adjustment). The bank claimed that this correction should reduce the earnings announced in the previous quarters, without specifying which. This implies that the decision is linked to credit problems that occurred during the 2008 crisis. The second loss concerns Discover Financial Services, which announced on February 12, 2010, that it would pay its former parent corporation Morgan Stanley $775 million to settle a breach of a contractual agreement. The case started in October 2008, when Morgan Stanley filed a complaint against Discover Financial Services concerning the distribution of proceeds from the resolution of antitrust litigation against rival issuers of Visa and MasterCard credit cards.

The third loss is $722 million. On November 4, 2009, the SEC announced a settlement whereby JP Morgan Securities paid a fine of $25 million to the SEC, and $50 million to Jefferson County, and dropped its claim for $647 million in termination fees linked to bonds and interest rate swaps. This settlement follows the sentencing of a former civil servant for accepting bribes. Originally, Jefferson County was verging on bankruptcy in February 2008. The $3 billion refinancing of its sewage system collapsed during the credit crisis. JPMorgan was the leader in banking transactions.

Fourth, in February 2010, the SEC and the Massachusetts authorities announced that the State Street Bank and Trust agreed to pay damages and fines under a judgment following allegations that the bank had misled some bonds investors about "Limited Duration Bond Fund" in 2007. The SEC also accused the bank of having provided information on these funds internally, which would have let some investors redeem the bonds early to the detriment of others who did not have this information. According to the SEC, the State Street Bank and Trust began to market the Limited Duration Bond Fund, which it described as "enhanced cash," in 2002. Many investors saw it as an interesting alternative to the money market. The problem was that in 2007, these funds were almost entirely invested in subprime residential securities and derivatives, which is much riskier than what the bank suggested in its communications.

For the fifth loss, according to the SEC, Bank of America omitted to accurately report to shareholders the losses on Merrill Lynch's books before the final ratification vote of the acquisition of Merrill Lynch. Bank of America was ordered to pay $150 million. The sixth and final loss occurred in September 2009: a businessman pled guilty and was sentenced to 12 years in prison for defrauding Bank of America ($142 million), Citigroup ($75 million) and HSBC ($75 million), a case of external fraud totaling more than $292 million. Apart from this case, the losses cited are linked to problems with information disclosure or errors related to risk management of financial products, particularly pricing, during the financial crisis. All of 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.

Further, credit risk always exists, and is highly influenced by Shadow Banking. Largely comprising false declarations and improper transactions, Shadow Banking is quite prominent in credit portfolios. Over $500 billion in credit "left" the banks' balance sheets and somehow transformed into Asset-Backed Commercial Papers between 2004 and 2007. This new way of skirting capital regulation, which bankers found too costly, reached a total of $1.3 trillion in July 2007 (Kroszner and Strahan, 2013; Acharya, Schnabl, and Suarez, 2013). Kindelberger and Aliber (2005) argue that "... as the monetary system gets stretched, institutions lose liquidity and unsuccessful swindles are about to be revealed, the temptation to take the money and run becomes virtually irresistible."

We now examine more losses from the 2008 crisis. Citigroup paid a total of $8.045 billion in March 2008 for the Enron scandal. Earlier, in October 2007, CFC lost $1.2 billion following the first waves of default in the subprime market. Bank of America intervened and ultimately bought out CFC. To continue this historical review, Goldman Sachs sustained a loss of $768 million in August 2008 concerning ARS (Auction Rate Securities). This bank was obliged to buy back 1.5 billion of these market instruments and paid penalties on this transaction. In another case of CliPBP, Bank of America had the same experience on a larger scale, and bought back 4.5 billion in ARS, for a total loss of $720.7 million in January 2009. OpVar categorizes the latter two losses as Trading and Sales business, which represents most CliPBP cases with Corporate Finance business.

In conclusion, in 90% of cases of operational losses, credit is pivotal to a history of improper transactions, along with Corporate Finance, Trading and Sales and/or Retail Banking. 80% of the amounts in question are attributable to two (Table 6) to six (Table 7) cases. In addition, it is often the same banks that are involved. Note that these historical spotlights were done by following “special” periods underscored by the regime shift detected. In other words, the regime detected seems to concern a set of banks in particular. We have documented 80% of the severity of operational losses by about only 20 cases, involving less than eight banks.

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
	Others (< 80%)	3.4 billion	30 losses			
	All	20.6 billion	32 losses			

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
	Others (< 80%)	753 million	21 losses			
	All	4.05 billion	27 losses			

3.1.5 Specification Test of the Hidden Markov Model

We now statistically test the validity of the HMM specification for our data. To do so, we follow 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 can be calculated with $z_t = \Phi^{-1}\left[Pr\left(y \leq y_t \middle| y_{t-1} = y_{t+1}\right)\right] \Rightarrow z_t \sim N(0,1)$. For details, we refer to Zucchini and MacDonald (2009).

Figure 6 shows the following points. The distribution of the monthly losses (in log) is asymmetrical (upper panel). The Skew t type 4 component is situated to the right of the mean to take this asymmetry into account (middle panel). The distribution of pseudo-residuals looks quite close to normal (bottom panel). This will be confirmed by the statistical tests. We now consider the statistical results 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 10% for this series of losses, as expected.

As for our model (pseudo-residuals), the Anderson-Darling test gives a p-value of 0.0682. This rejects normality even if this p-value is not far from 10%. Conversely, 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. This seems to show that 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.

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

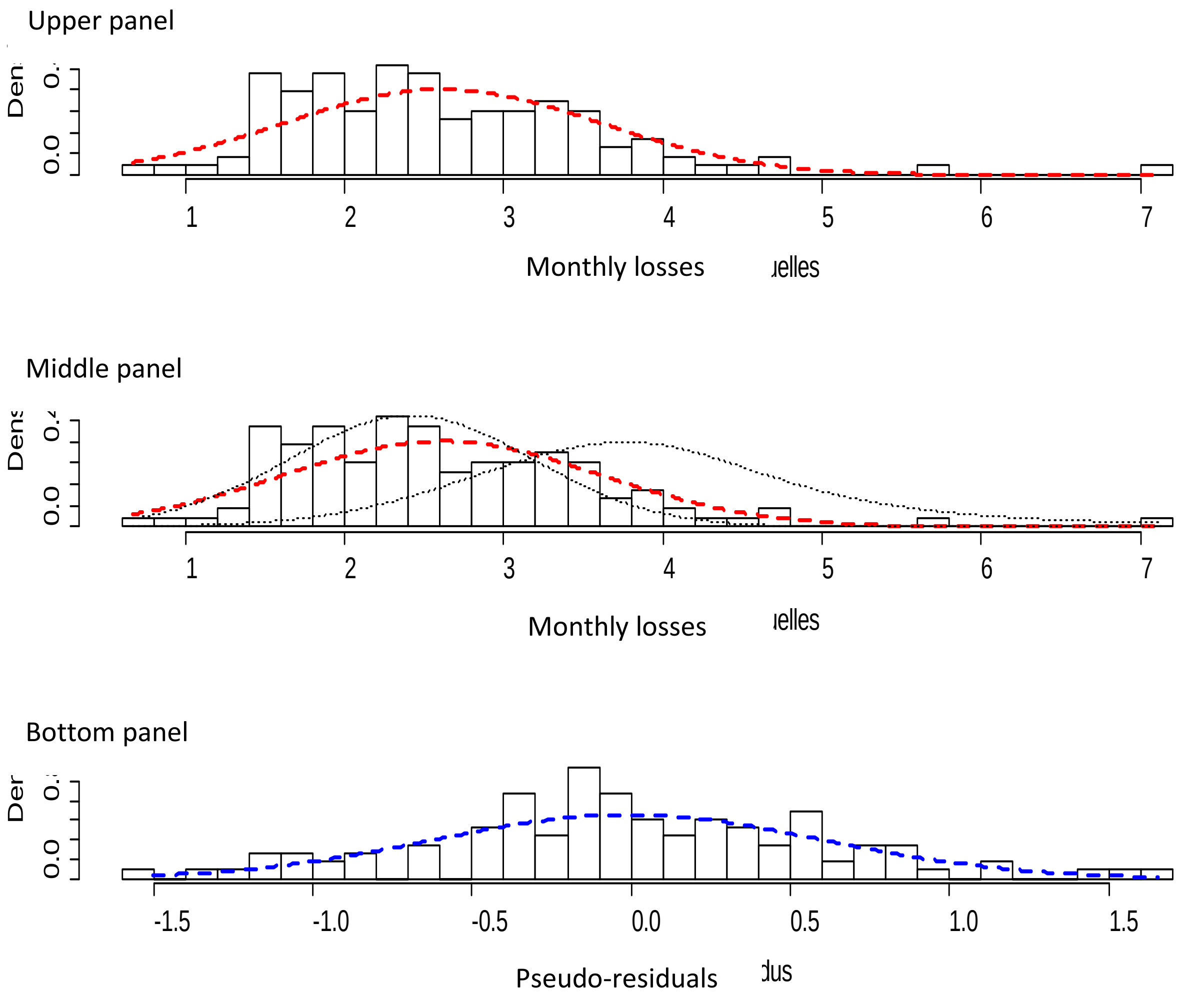

Figure 6: Histograms of monthly losses and pseudo-residuals

4 Measuring the effect of regimes detected

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

$$log(Loss) = \alpha + \beta\, log(Assets) + \lambda BusinessLines + \delta EventTypes + \in . \qquad (4.1)$$

The dependent variable is *log(Loss)*. The independent variables are *log(Assets)*, category variables Business Lines, *BL*, and category variables EventTypes, *ET*. The fixed time effects are years.

The regressions results are presented in Table 9. Model (1a) is the reference model. To simplify the presentation of the estimates, we do not report the coefficients of the year fixed effects (Year FE), because they are not pertinent to the discussion. A "yes/no" indication for their presence is presented in the table. We add the variable of the HMM regime only in model (2a) and its cross-loadings (interaction) with Business Lines and Event Types in (3a). [1] 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)*

	(1a) Reference model	(2a) Adding HMM regime	(3a) Adding HMM regime and crossings
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 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)

[1] The model has also been estimated using Heckman's model to consider potential endogeneity of firms that sustained losses, as in Dahen and Dionne (2010). The results are available from the authors. They indicate that the inverse Mills ratio is not significant in the second step; the other results remain comparable to those in Table 9.

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. Mnmt			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 FE	yes	yes	yes
Adj. R^2	0.170	0.186	0.223
AIC	1993.5	1985.2	1978.04
Log Likelihood p-value Chi2	-971.8	-966.6 0.001 (2a vs 1a)	-952.0 0.002 (3a vs 2a)
Num. obs.	508	508	508

Note: *** p < 0.01, ** p < 0.05, * p < 0.1. Clients products and business practice and retail brokerage are the omitted categories for Event Types and Business Lines, respectively.

The variable *log(Assets)* is very significant, which is consistent with this type of model. The coefficient tends to keep the same magnitude in all regressions. The coefficient of the high regime variable is very significant at 1% in model 2a but less significant in model 3a, where it is significant 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 2a or cross-loaded in 3a. The AIC statistic and the Log Likelihood ratio test also confirm the superiority of model 3a. That being said, we must perform backtesting on these models to evaluate their validity and calculate the reserve capital. Note that in the loss database there were no 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 3a.

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 (2010). Let Y be a random variable that follows a negative binomial law with average λ and the dispersion parameter θ. If f_{NB} is the probability mass function of this law, then the probability that Y is equal to a value *k* is written as:

$$Pr\left(Y=k|\lambda,\theta\right)=f_{NB}\left(k,\lambda,\theta\right)=\frac{\Gamma\left(k+1/\theta\right)}{k!\Gamma\left(1/\theta\right)}\left[\frac{1}{1+\theta\lambda}\right]^{\frac{1}{\theta}}\left[\frac{\theta\lambda}{1+\theta\lambda}\right] \quad (4.2)$$

where $k=0,1,2,...,\Gamma(.)$ designates the conventional gamma function. Note that $\theta>0$ and that the negative binomial converges toward a Poisson law when $\theta\to0$ (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 negative zero inflated 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:

$$Pr\left(Y_{ij}=k\right)=\begin{cases}q_{ij}+\left(1-q_{ij}\right)f_{NB}\left(0,\lambda_{ij},\theta\right) & k=0\\ \left(1-q_{ij}\right)f_{NB}\left(k,\lambda_{ij},\theta\right) & k=1,2,...\end{cases} \quad (4.3)$$

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

$$\begin{aligned} log\left(\lambda_{ij}\right) = \zeta_0 + \zeta_1 log\left(Assets_{ij}\right) + \zeta_2 RegimeHMM + \zeta_3 GDP \\ + \zeta_4 Bank\_Cap_{ij} + \zeta_5 Mean\_Salary_{ij} \end{aligned} \tag{4.4}$$

$$log\left(\frac{q_{ij}}{1-q_{ij}}\right) = \xi_0 + \xi_1 log\left(Assets_{ij}\right) + \xi_2 RegimeHMM + \xi_3 GDP + \xi_4 Mean\_Salary_{ij}. \tag{4.5}$$

The last formula is equivalent to the modeling of q_{ij} using the logistic distribution. The variable *log(Assets)* is the total assets of the bank (in log) and the variable HMM is for the High Regime. Mean-Salary is the mean salary paid in the bank, Bank_Cap is the bank capitalization and GDP is Gross Domestic Product during the period.

The estimates are presented in Table 10. The dependent variable is the number of annual losses. In (1b) we present the benchmark model to compare the effect of adding regimes: 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 (2b) adds this dimension in both parts. Its 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. We want to measure whether deflation of zeros provides statistical value. To do so, we compare this deflation model with the base model 1b. Knowing that they are embedded, we can test it with the likelihood ratio whose results appear below in the same table. The likelihood ratio test of model 2b versus 1b is conclusive, with a statistic of 46.53 and a p-value of almost 0. Model 2b using the Markov regime seems to provide more information than the reference model (1b) 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 theta dispersion parameter of the negative deflated binomial model. Starting with a value of 2.097 in model 1b, we reach 1.085 for 2b, which is a clear improvement in the specification in the sense that there is less unobserved heterogeneity in 2b. We can proceed to the backtesting of the model.

Table 10: Effect of regimes on frequencies

	(1b) Reference model	(2b) 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(theta)	2.097***	1.085***
	(0.634)	(0.417)
Zero 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: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5 Backtesting

5.1 Operational loss capital

This section has a dual objective. First we want to construct a backtesting procedure for our models with regimes to determine their validity. We also want to measure the extent that ignoring the existence of regimes in our operational loss data biases calculation of reserve capital if this reality is not formally considered. The period selected to calculate coverage is January 2010 to December 2010. This period will be designated by *Couv*0. The regime is high for the month of January and low for the 11 other months. We number our three models as follows: #1 base model; #2 Markov regime; #3 Markov regime + cross-loading with Business Lines and Event Types. To extend Dahen and Dionne (2010), we construct our backtesting by taking into account regimes detected. There will be an In-Sample backtesting calculation, in the sense that the history will include the period *Couv*: from January 2001 to December 2010 (called *Hist*1). Further, by definition, Out-of-Sample backtesting does not include the period covered in the history, and will last from January 2001 to December 2009 (designated by *Hist*2). For each model, the data from the periods *Hist*1, *Hist*2 and *Couv* 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 the period *Couv*. Once scaled for a given bank, the historical losses (*Hist*1 or *Hist*2) can be considered to follow a lognormal distribution. If we consider the bank U.S. Bancorp (Table 11), 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 three 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 performed 200,000 observations from the lognormal in question, for which we calculate the convolution for 2,000 numbers drawn from the negative binomial of the corresponding frequency model. This gives us a distribution for which we calculate the reserve capital 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 in the case where the VaR at 99.9% is not exceeded.

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 number of such values; and lastly the Christoffersen (1998) test, which helps us determine the conditional simultaneous coverage of frequency and independence of the values in excess of VaR. This gives us a complete and robust view of the validity of our backtesting. To provide figures, we have 445 losses recorded for the period *Hist*1 and 63 for the period Cov, which gives us 508 = 445 + 63 losses for *Hist*2. We must calculate the probable losses that a given bank incurs during the period *Couv*. To do so, the 63 losses of *Couv* 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. Note that the scaling will cover all historical losses of *Hist*1 (in-sample) or *Hist*2 (out-of-sample) and all possible losses during the period *Couv*. Consequently, the model that passes backtesting is automatically that which successfully allows simultaneous scaling of all the loss observations in question.

We perform the calculations for two banks. The first is U.S. Bancorp (as in Dahen and Dionne, 2010). Table 11 indicates that the Kupiec test rejects the VaR at 95% in in-sample for base model #1 (no regime). The reason for this is that the excess values observed are too few, at 3.4% versus 5% theoretical. For the rest of the degrees of confidence of model #1 for in-sample and out-of-sample, all seems to function properly. The same pattern is seen regarding independence of the values in excess of VaR except for the VaR at 99.5% in out-of-sample, where the DQ test rejects the validity at 5%, whereas the Christoffersen test still does not allow us to reject it at 5%. Capital at 99.9% is $2,957.4 million. The bank's total assets are $290.6 billion, and reserve capital represents 1.02% of assets. Model #2 shows a weakness in the frequency of values in excess of VaR at 95% and 99.5% in in-sample, and VaR at 95% in out-of-sample. We observe the same weaknesses in model #3 concerning VaR at 95%, 99.5% in in-sample, and 99% in out-of-sample. For the independence of draws, the DQ test is rejected at 5% for VaR at 99.5% in-sample, and all else is correct at 5%. The Christoffersen test shows the same weakness in in-sample for VaR at 5% and at 99.5%, and the rest is correct at 5%. Concerning the reserve capital calculated, it is lower than for benchmark model #1, with $2,480.5 million and $2,060.7 for VaR at 99.9% in model #2 and model #3 respectively.

We conclude with two important remarks. The first is that all capital calculated is below that calculated for model #1, which does not take into account the existence of regimes. This finding supports what we said at the beginning of the paper: that there is an endogenous Hidden Markov regime in our data and that ignoring it amounts to injecting a positive bias to calculate capital when the regime is at a low level. Conversely, a negative bias increases the risk of underestimating the reserve capital required when the regime level is high. Using the calculation of model #3, this bias for U.S. Bancorp is (2957.4-2060.7)/2957.4, which is 30.3% too high. The second comment is that the various weaknesses shown by the tests above seem to mainly arise in VaR at 95%, and always concern excess (very high) reserves. We thus consider that models #2 and #3 are validated by backtesting. In addition, model #3 stands out from the others by allowing considerable savings in capital.

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 #3 is still the least capital expensive. Note this time that models #2 and #3 do not pass the Kupiec test in out-of-sample at 99.9%. The same comment can be made for the DQ and Christoffersen tests. However, VaR at the intermediate level of 99.5% seems to respond well in the same tests. Note that model #1 is also at the limit of rejection at 5% for the same VaR at 99.9% in out-of-sample with a p-value of 0.0506. If we consider model #3 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. Consideration of Markov regimes thus provides an irrefutable improvement.

5.2 Number of states in HMM model

To further backtest own research, we raise two questions. The first would be whether we can statistically justify that a combination of two normals, instead of one normal and an ST4, would have been insufficient. The second question is to ask 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 want to compare our model N+ST4 to two other models: 2N and 2N+ST4. The estimates imply that we would not have a better specification than N+ST4. We tested the normality of the pseudo-residuals of the three models as shown in Table 13. First, concerning the model 2N with two levels, all three p-values are below 10%. The data clearly show that this model is not adequate. 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 reason at 10%, we have two tests that reject normality whereas only Anderson-Darling showed a problem for the two-level N+ST4, as seen above. In addition, the value of the AIC criterion of the model 2N+ST4 is 325.59 versus 321.93 for our two-level model N+ST4, which indicates deterioration in performance. This deterioration is more evident when we use the criterion BIC, which becomes 380.66 for the three-level, whereas it was 352.22 for the model N+ST4. We therefore reject the three-level model 2N+ST4 at a level of confidence of 10%. Consequently, we definitively retain the two level specifications with a normal law and one Skew t type 4 for our extreme observations.

Another comment is necessary. A priori, if the data allow a sufficient number of observations and quality, we should have a better goodness of-fit if we increase the degrees of freedom of a given model. In our case, according to Figure 3.6, there are 18 observations representing high loss regime. The addition of a third level would have divided up these 18 observations into two levels. The three resulting levels would be "normal losses," "large losses" and "very large losses." However, the 18 observations are too few to model two distinct levels. In addition, very few periods start from the ST4 level, which makes this level non-significant. Lastly, in this case it is as if we had a first level represented by a normal, followed by a second with a second normal. This three-level model is therefore effectively reduced to two-level regime with two normals only, because the ST4 level is not representative. Hence the p-values of the three-level regime let us reject the three-level model, together with the two-level model built with two normal distributions.

Table 11: Backtesting of U.S. Bancorp bank

Backtesting	Model	α (Frequency)		VaR	Kupiec test		DQ of E-M		Christoffersen	
Reference model		Theoretical	Observed		Stat.	p.value	Stat.	p.value	Stat.	p.value
In-Sample	1	0.050	0.034	269.7	11.039	0.0009	10.246	0.0365	11.797	0.0027
	1	0.010	0.009	842.3	0.233	0.6292	5.129	0.2744	0.233	0.8899
	1	0.005	0.004	1289.7	0.116	0.7334	0.208	0.9949	0.116	0.9436
	1	0.001	0.002	2957.4	1.991	0.1583	2.751	0.6063	1.991	0.3696
Out-of-Sample	1	0.050	0.043	269.7	1.760	0.1846	2.359	0.6701	2.550	0.2795
	1	0.010	0.012	842.3	0.479	0.4887	1.404	0.8434	0.479	0.7869
	1	0.005	0.008	1289.7	2.385	0.1225	14.792	0.0052	5.027	0.0810
	1	0.001	0.002	2957.4	1.991	0.1583	2.751	0.6003	1.991	0.3696
HMM regimes										
In-Sample	2	0.050	0.036	230.1	8.755	0.0031	9.183	0.0567	8.812	0.0122
	2	0.010	0.010	712.1	0.000	1.0000	4.229	0.3759	1.775	0.4118
	2	0.005	0.009	1067.3	5.659	0.0174	8.089	0.0884	5.659	0.0590
	2	0.001	0.002	2480.5	0.666	0.4145	0.826	0.9349	0.666	0.7169
Out-of-Sample	2	0.050	0.039	230.1	4.538	0.0332	5.253	0.2623	4.541	0.1033
	2	0.010	0.012	712.1	0.479	0.4887	3.802	0.4334	0.479	0.7869
	2	0.005	0.004	1067.3	0.484	0.4867	0.511	0.9724	0.484	0.7851
	2	0.001	0.002	2480.5	1.991	0.1583	2.751	0.6063	1.991	0.3696
HMM regimes and interactions										
In-Sample	3	0.050	0.035	209.0	9.483	0.0021	8.768	0.0672	9.520	0.0086
	3	0.010	0.011	619.3	0.217	0.6416	6.884	0.1421	1.653	0.4376
	3	0.005	0.010	913.6	6.999	0.0082	10.165	0.0377	6.999	0.0302
	3	0.001	0.001	2060.7	0.425	0.5146	0.357	0.9859	0.425	0.8086
Out-of-Sample	3	0.050	0.043	209.0	1.760	0.1846	3.430	0.4886	1.844	0.3977
	3	0.010	0.016	619.3	5.730	0.0167	9.258	0.0550	5.730	0.0570
	3	0.005	0.004	913.6	0.116	0.7334	0.208	0.9949	0.116	0.9436
	3	0.001	0.002	2060.7	0.666	0.4145	0.826	0.9349	0.666	0.7169

Table 12: Backtesting of Fifth Third Bancorp bank

Backtesting	Model	α (Frequency)		VaR	Kupiec test		DQ of E-M		Christoffersen	
Reference model		Theoretical	Observed		Stat.	p.value	Stat.	p.value	Stat.	p.value
In-Sample	1	0.050	0.028	115.0	20.948	0.0000	18.955	0.0008	21.130	0.000
	1	0.010	0.008	430.7	0.972	0.3241	1.161	0.8844	0.972	0.6150
	1	0.005	0.007	689.8	0.909	0.3403	1.334	0.8556	0.909	0.6346
	1	0.001	0.002	1722.6	1.991	0.1583	2.751	0.6063	1.991	0.3696
Out-of-Sample	1	0.050	0.038	115.0	5.590	0.0181	7.547	0.1097	5.816	0.0546
	1	0.010	0.007	430.7	1.553	0.2127	1.614	0.8063	1.553	0.4600
	1	0.005	0.004	689.8	0.484	0.4867	0.511	0.9724	0.484	0.7851
	1	0.001	0.003	1722.6	3.822	0.0506	5.812	0.2137	3.822	0.1479
HMM regimes										
In-Sample	2	0.050	0.032	100.4	13.629	0.0002	17.235	0.0017	17.030	0.0002
	2	0.010	0.008	377.0	0.972	0.3241	1.161	0.8844	0.972	0.6150
	2	0.005	0.004	592.2	0.484	0.4867	0.511	0.9724	0.484	0.7851
	2	0.001	0.002	1522.9	0.666	0.4145	0.826	0.9349	0.666	0.7169
Out-of-Sample	2	0.050	0.042	100.4	2.415	0.1202	4.089	0.3941	2.536	0.2814
	2	0.010	0.010	377.0	0.000	1.0000	0.564	0.9670	0.000	1.0000
	2	0.005	0.009	592.2	4.439	0.0351	6.256	0.1808	4.439	0.1087
	2	0.001	0.005	1522.9	14.599	0.0001	29.515	0.0000	14.599	0.0007
HMM regimes and interactions										
In-Sample	3	0.050	0.031	94.4	15.532	0.0001	16.867	0.0021	17.537	0.0002
	3	0.010	0.008	338.1	0.972	0.3241	1.161	0.8844	0.972	0.6150
	3	0.005	0.004	522.6	0.484	0.4867	0.511	0.9724	0.484	0.7851
	3	0.001	0.002	1291.5	0.666	0.4145	0.826	0.9349	0.666	0.7169
Out-of-Sample	3	0.050	0.042	94.4	2.783	0.0953	4.984	0.2889	2.804	0.2461
	3	0.010	0.013	338.1	1.829	0.1762	3.369	0.4980	1.829	0.4007
	3	0.005	0.007	522.6	1.570	0.2102	2.205	0.6981	1.570	0.4562
	3	0.001	0.003	1291.5	6.057	0.0138	10.012	0.0402	6.057	0.0484

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

		(1) 3-level model 2 Normals + 1 ST4		(2) 2-level model 1 Normal + 1 ST4		(3) 2-level model 2 Normals	
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

Conclusion

In this article, we analyze the effect of business cycles in operational loss data on optimal capital of banks. We show that considering business cycles can reduce capital for operational risk by redistributing it between high regime and low regime states. The variation of capital is estimated to be in the range of 25% to 30% in our period of analysis. We also demonstrate 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 is also observed for significant losses sustained by insurance companies whose settlement payments are often determined by the courts.

Several extensions of our study are possible. The most promising would be to verify the stability of the results using different regime detection methods (Maalaoui, Chun et al., 2014). How can an approach to detect regimes in real time improve the results, and in particular take the asymmetry detected in this article into account? The value of this approach is that it allows separate analysis of level and volatility regimes.

Another possible extension is to use a different approach than that of scaling of operational losses to generate a larger number of observations at each bank. Some banks use the Change of Measure Approach proposed by Dutta and Babbet (2013). This method combines scenario analysis with historical loss data. It would be interesting to examine whether the results of this approach can remain stable by introducing cycles in the data. It would also be worth extending the analyses to stress testing of models.

References

Acharya, Viral V, Philipp Schnabl and Gustavo Suarez (2013). Securitization without risk transfer. *Journal of Financial Economics* 107(3), 515-536.

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

Baum, Leonard E and Ted Petrie (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics* 37(6), 1554-1563.

Baum, Leonard E, Ted Petrie, George Soules and Norman Weiss (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics* 41, 164-171.

Bulla, Jan (2011). Hidden Markov models with *t* components. Increased persistence and other aspects. *Quantitative Finance* 11(3), 459-475.

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

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

Dahen, Hela, Georges Dionne and Daniel Zajdenweber (2010). A practical application of extreme value theory to operational risk in banks. *Journal of Operational Risk* 5(2), 63-78.

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

Dutta, Kabir K and David F Babbel (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, Robert F and Simone Manganelli (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* 22(4), 367-381.

Hamilton, James D (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57(2), 357-384.

Kindelberger, Charles P and Robert Z Aliber (2011). *Manias, panics and crashes: A history of financial crisis*. Palgrave Macmillan, 356 p.

Korolkiewicz, Malgorzata W and Robert J Elliott (2007). *Smoothed Parameter Estimation for a Hidden Markov Chain of Credit Quality*. In: RS Mamon and RJ Elliott (Eds), *Hidden Markov Models in Finance*. Springer US, pp. 69-90.

Kroszner, Randall S and Philip E Strahan (2013). Regulation and deregulation of the US banking industry: Causes, consequences and implications for the future. In: Nancy L Rose (Ed.) *Economic Regulation and Its Reform: What Have We Learned?* University of Chicago Press, pp 485-543.

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

Liechty, John (2013). Regime switching models and risk measurement tools. In: Jean-Pierre Fouque and Joseph A Langsam (Eds): *Handbook on Systemic Risk*. Cambridge University Press, pp. 180-192.

Maalaoui Chun, Olfa, Georges Dionne and Pascal Francois (2014). Detecting Regime Shifts in Credit Spreads, *Journal of Financial and Quantitative Analysis* 49(5-6), 1339-1364.

Mitra, Sovan and Paresh Date (2010). Regime switching volatility calibration by the Baum-Welch method. *Journal of Computational and Applied Mathematics* 234(12), 3243-3260.

Nešlehová, Johanna, Paul Embrechts and Valérie Chavez-Demoulin (2006). Finite mean models and the LDA for operational risk. *Journal of Operational Risk* 1(1), 3-25.

Psaradakis, Zacharias and Martin Sola (1998). Finite-sample properties of the maximum likelihood estimator in autoregressive models with Markov switching. *Journal of Econometrics* 86(2), 369-386.

Rabiner, Lawrence (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2), 257-286.

Rigby, Bob, Mikis Stasinopoulos, Gillian Heller and Vlasios Voudouris (2014). The Distribution Toolbox of GAMLSS. http://www.gamlss.org/wp-content/uploads/2014/10/distributions.pdf.

Siu, Kin Bong and Hailiang Yang (2007), Expected shortfall under a model with market and credit risks. In: RS Mamon and RJ Elliott (Eds.) *Hidden Markov Models in Finance*. Springer US, pp. 91-100.

Siu, Tak Kuen (2007). On fair valuation of participating life insurance policies with regime switching. In: RS Mamon and RJ Elliott (Eds.) *Hidden Markov Models in Finance*. Springer US, pp. 31-43.

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

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