A markov model of interbank dynamics with adaptive techniques

Duc PHAM-HI

Head of Financial Engineering Dept.,
ECE Paris graduate school of Engineering, 37 quai de Grenelle,75015 Paris
phamhi@ece.fr

ABSTRACT: *This paper will first provide a critique of the modeling practices currently in use in most banks with the agreement of the national supervisors in various implementations of Basel II regulations. It then suggest a set of stochastic equations to make models more time dependent and more causally risk sensitive. Following a discussion explaining the abandon of the search for optimization solutions through the resulting Hamilton-Jacobi-Bellman equation, we expose a toy implementation comprising a commercial bank sector lending to the public, buying protection from other banks and receiving "quantitative easing" from a central bank.The amount of quantitative easing is derived from a learning process, using filtering techniques. Filiation from online sampling and learning justify of the use of such techniques.*

KEY WORDS: Banking risks, central bank intervention, Risk management modeling, Levy process, HJB equation, Particle filters, Adaptive control.

1. INTRODUCTION

Risk management is not primarily about statistics of losses. It is about adopting coherent and optimal policies to face potential disaster. It is also about drawing from a global vision of your organisation, and your environment, and the means at your disposal, to derive intelligent behaviour while facing risks. Behaviour will be defined here as a set of actions linked together by the same rationale, motivations or preferences, either economic or psychological. Taking into account the deficiencies in Basel II, some of which Basel III is trying to mend, we shall adopt an interbank modeling framework, and embed in it the flow of time.

Thus, in managing banking system risk, and leaving aside the market risks with its internal fraud losses, we will focus on questions such as:

- How do banks react when they face large, own-funds shrinking, operational or credit losses?

- How does the central bank learn how much to inject in terms of liquidity to reduce the risk of bank collapse?
- What are the main parameters of such a model?

The answer is tentatively given in the implementation of a toy model that will link the search of an online learning process to intervene in the face of bank losses. We have to stress this is still work in progress and the model still need much tuning.

1. LACK OF TEMPORAL DYNAMICS IN RISK MANAGEMENT PROCESSES

Today's Risk management models used in banks are, with the general exception of market risks models, only static models, as opposed to dynamic models. Indeed, unlike dynamic hedging in portfolio asset management, Credit risk and Operational risks

In market risks, one is supposed to watch for the 10-days Value-at-risk, so it is very close to being time sensitive. Unfortunately, it is more like a sliding time-frame than a chain of events. Dynamically hedging portfolios is more time-based, but it is more a earnings related concern than a risk management concept.

In credit risks, Basel 2.5 has introduced IRC (Incremental risk charge) which was originally a Credit notion only. However, it has failed to be generalized to the rest of risk management. Asymptotic single risk factor (ASRF) seems to be the only notion where sensitivity of risk to many temporal variables and parameters is formally framed. However, it serves as a justification to a static formula that states the quantity of capital needed for credit risk.

The status in operational risks is even more static. The current predominant model, Lost Distribution Approach, by supervisory recommendation, should be based on 5 years data. Most banks do a yearly campaign of risk scenarios collection; therefore their computation of operational Value-at-risk has a periodicity of one year. In most cases, they add the new data to the stock of historic data. Therefore, parameter evolutions will suffer from a large and increasing inertia.

Furthermore, the only indicators that play a role on the risk dashboard, namely the business environment and internal control factors (BEICF), are rather « status indicators » and not level indicators: they are qualitative in nature and are extremely difficult to quantify correctly and with a fine, sensitive granularity. In most cases, they are just "green-yellow-red" lights.

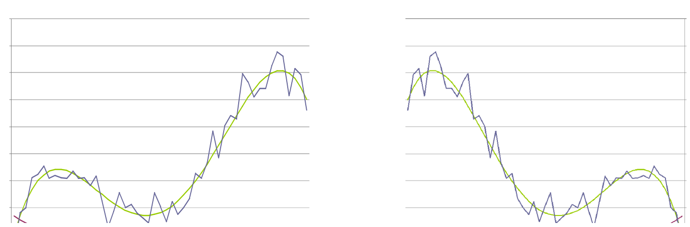


Fig. 1: These 2 banks with different risk histories have the same Value at risk

One of the consequences of this calculation in operational risk is that within the 5 years loss data history, if one is to reverse the scale of time , or even by changing the dates of loss occurrence, the losses are put in increasing or decreasing order, instead of chronologically, nothing will be changed about the Value-at-risk. In fact, it would be the same if all the losses over these 5 years had happened simultaneously, the first or last day of the period.

Another important motivation is to decompose risk into components so they can be calculated separately instead of as a lump sum. The advantage of separate calculations is in the finer granularity of causal factors. This allows further analytical judgments and tactical decision making.

Another consequence of modeling into components is the capture of each heterogeneous behavior and identifies the recurrent causes from the singular events. This in turn helps analyze tail events, which are, by definition, single large events.

The main drawback to this decomposition approach is the aggregation problems, including correlations estimation.

2. A DYNAMIC CONTROL MODEL FOR RISK MANAGEMENT

2.1. Introduction of time based processes of profit and loss

A quick analysis of frequency of occurrence versus size of loss reveals there are 2 major risk categories worthy of consideration: frequent but small to medium losses, and rare but catastrophic losses.

On one hand, small losses frequently occur in banking processing, due for instance to inevitable complexity errors, like mistyping a check's amount, causing its cost of rejection or everyday small creditors defaulting on small amounts. They vary with business volume or increase in business activity, and are "regular" enough for their *return* to vary in a Gaussian way. The *return* of "losses" can be either positive or negative. For this first risk category, between time t and $t+\Delta t$, the increment of loss is proportional to the current business volume, and to a degree of variability which is a sort of volatility as in:

$$dX_t^A = X_t^A \sigma_t dB_t$$

On the other hand, very rare but catastrophic events can be modelled as independent events, each having a constant probability of occurring in time, and causing losses according to a time-invariant probability distribution.

To handle that second risk category, it can be supposed that:

(i) Probability of catastrophes do not vary over time in current affairs (there is no reason to suppose that earthquakes , or viral attacks have an inherent time-dependent structure) : the lapse of time between two catastrophes follow the same probability distribution law in the past, present and future.

(ii) Catastrophes are not dependent on each other.

With these common sense assumptions, these "stationary increments and independent increments" observations constitute a big step in modeling these second category of risks. Indeed it comes as no surprise that for ages, insurers have relied upon a Poisson law distribution (in which time need not flow uniformly) to represent the number of incoming claims N as a function of time t and compound it with an independent loss size distribution $\{(z_k)\}$.

The total second category losses up to time t can thus be written:

$$X_t^B = \sum_{j=1}^{N(t)} x_j = \sum_{k=1}^{d} z_k N_t^k$$

where the $(z_k)_{k=1,\ldots,d}$ are different non-null numbers, the d Poisson processes $(N^k)_{k=1,\ldots,d}$ are independent.

In fact, this lies at the core of most "Loss Distribution Approach" advanced models considered in the New Basle Accord's approach for sizing up requirements for operational risk capital. In Credit risk we can lump large, closely correlated defaults, into one catastrophic portfolio loss.

In differential form,

$$dX_t^B = \beta X_{t-}dt + X_{t-}\int_{\mathbf{R}^*}\left(e^z - 1\right)\widetilde{N}(dt,ds)$$

where $\widetilde{N}(dt,dz) = N(dt,dz) - dt \times \nu(dz)$ is the compensated Poisson random measure [9]. The deterministic measure $\nu(dz)$ is the Lévy measure, and $dt \times \nu(dz)$ designates the intensity of the jumps between t and t+dt with the technical condition $\int_{\mathbf{R}}\left|e^z - 1\right|\nu(dz) < \infty$ under the compound Poisson process. This intensity will combine with a random severity of loss process that can be specified later using collected databases of loss history to be fitted to observations.

We can now synthesize a loss function taking advantage of the Lévy-Khintchine representation. The sum of these two different types of losses yields a single Lévy process. In this general process, both the fat tails well described by the Pareto distributions in Extreme Value Theory, and the business generated “expected losses” described in the Basel terminology, coexist. Collecting these terms, the following description of the total loss is obtained as

$$dX_t = dX_t^A + dX_t^B.$$

The loss process, as a combination of pure jumps Poisson process with a Brownian motion with drift, thus follows a Levy process whose characteristic triplet is given by $(\beta(=0), \sigma^2, \nu(dz))$. Its representation can be expressed with the Lévy-Khintchine decomposition:

$$dX_t = \sigma_t X_t dB_t + X_{t-}\int_{\mathbf{R}^*}\left(e^z - 1\right)\widetilde{N}(dt,ds)$$

2.2. Introducing Management and Environment qualitative variables

These losses can be reduced by better quality management, and by raising risk awareness and reporting through ad hoc budgets. By taking risk budget as a control variable, quality of management is naturally introduced into the wealth evolution equation.

We will describe here two effects:

- The volatility of the Gaussian component of everyday medium risk, as well as the size of jumps in the Poisson component of catastrophic risk, can be subordinated to mitigation effects arising from better quality management and more proactive risk management.
- The industry's so-called "key risk indicators", and the scorecards approaches to risk reporting, provide early warning systems that can reduce the impact of losses. They also provide a measure of the ambient, environmental risk intensity at a particular moment in time.

Their effects can be modeled as coefficients modulating the amplitude of the draw-downs. This effect is introduced here through control variables that represent budget amounts ζ and ξ per unit time [1] that Management has levied from the income per unit time μ to spend on risk prevention actions. Qualitative assessments such as scorecards techniques, give a feedback on the perceived level of residual risk. Intuitively, quality of risk management should introduce a differentiation effect on the required capital: a more proactive mitigation of risk should require less capital than no action is taken to reduce risks.

Spending and investing ζ euros in a scorecard or another risk measure approach should reduce the variance of small, frequent risks by a factor of η, so that residual risk is now down to $\sigma_t \eta(\zeta)$ from σ_t .However, the transformation of this expense into a risk mitigant cannot have linear effects indefinitely ; therefore we add conditions : $0 < k_1 < \eta \leq 1$ and $\eta'(\zeta) \leq 0$.

Reduced risks on expected losses can thus be expressed as:

$$X_t^A = X_0 \exp \sigma_t \eta(\zeta) B_t$$

It is commonly observed that there are some periods where business activity is more risky than at others or, in some places more than others. Examples are economic downturns, atmospheric or geologic hazard (typhoons zone or period of airline controller strikes…). Let us introduce a parameter θ that will stand for a business environment indicator, a "Key Risk Indicator", a sort of "temperature" of "hot" or "cool" atmosphere to do business or credit. Also, θ should increase, ceteris paribus, if business is expanding aggressively: thus, $\theta'(\mu) \geq 0$ (negative effect of agressive market strategies or expanding business on operational risk). This effect then increases or decreases the losses from value $\sigma_t \eta(\zeta)$ to $\sigma_t \eta(\zeta) \theta(\mu)$.

We now turn to the effects of contingency planning on catastrophic risks. The impact of risk budget cannot always be readily assessed, because some of the threats being hedged against will occur once in a thousand years. Here, utility function showing aversion to risk can be used to create a component of the value

---

[1] or sequences of $\{\zeta_t, \xi_t\}$ in [0, T]. We drop the time t subscript for ease of reading.

function that will be sensitive to another type of budget control variable. The hypothesis on the impact of insurance or recovery plans, costing a budgeted amount ξ, covering catastrophic events, is that these measures reduce losses by a coefficient g which is a function of ξ.

Residual risk decreases to :

$$F_{\xi}(X_t^B) = \sum_0^{N(t)} K_{\xi}(x_j^B) = g(\xi)\int_{\mathbf{R}+}(e^z - 1)\tilde{N}_t(dx)$$

with $k_2 < g(\xi) < 1 \quad g'(\xi) \leq 0 \quad g'' \geq 0$ where individual losses x_j may be capped or reduced, where k_2 is a constant representing the insurance waiver level (CDS franchise).

The Revenue process is supposed here to be deterministic; It generates an income at rate μ_τ, and is proportional to total cash available for the period, both for investing into risk prevention, and for raising wealth level. In the toy model presented below, it is composed of the interests received on loans to individual creditors and premium collected over CDS (protection from default) sold to other banks, minus the premium paid on protection bought.

$$dR_t = \mu_t W_t dt$$

Wealth for the period is made up of losses and gains, and putting $dW_t = dR_t + dX_t$ we get its evolution equation:

$$dW_t = W_t(\mu_t - (\varsigma + \xi))dt - W_t\sigma_t\eta_{\varsigma}\theta\, dB_t - g(\xi)W_t\int_{\mathbf{R}}(e^z - 1)\tilde{N}_t(dx)$$

It can be remarked here that dW_t can be allowed to be (temporarily) negative, provided W_t stays positive. The technical condition to add here is that the period under consideration is limited to $[0,T_f]$ where T_f is the stopping time, at the first bankruptcy event. The cumulative level of cash up to time t is given by

$$W_t^{\pi} = \int_0^t dW(s, \mu_t, \sigma_t, \pi_t, N, \nu)$$

where the dependency of cash on policy $\pi(t)$ as the given pair ($\zeta(t)$, $\xi(t)$), is highlighted, both through the expense that cuts into business income and the savings from losses. It is of course also dependent on the parameters of the two types of risks, N and ν .

2.3. Introduction of behaviour in risk management processes

We want to observe behaviour of banks, given the rationality in the gains and losses in the processes exposed in the previous section.

In modeling optimal behaviour, we have to introduce actions of risk management. At time t, let a_t be one possible action in a set of possible choices while facing a risk situation characterized by state variable x_t in a probabilised set Ω .

$$a \in \pi = \{a_t\} = \{a_t(x_t)\}$$

Choosing a set a will result in a "reward" (or penalty) R, which in a probabilistic context, is the expectation of a system's response r as a function of x and a .

$$R(x,a) = E\{r(x,a)\}$$

Reward has to do with Value, as it increments the riches held in the system. A risk policy π is taken as one time- and state- dependent set of possible actions. Supposing null terminal value at infinite t, the value function is the total of what can be expected in the future, introducing a discount rate γ and taking the expected value

$$V^{\pi}(x) = E\left[\sum_{t=0}^{\infty} \gamma^t r(x_t, \pi(x_t)) \middle| x_0 = x\right]$$

A value V^{π} for the global period is thus associated to each risk policy π.

Rationality of the risk center is to seek maximum of value, starting from state x_0 at time t, to "learn" policy π^* maximising V over set A of admissible actions satisfying

$$V^{\pi^*}(x_0) = \min_A \left[V^{\pi}(x_0)\right]$$

where $V^{\pi}(x)$ is the the consequence of following policy π from initial situation x

$$V_{t+1}(x) = R(x, \pi_{t+1}) + \gamma . \sum_y P_{x,y}.V_t(y)$$

Value for a given strategy is sum of immediate reward and discounted flow of possible future rewards, depending on the transition:

$$V^{\pi}(x) = R(x, \pi(x)) + E\left\{\sum_{t=0}^{\infty} \gamma^t r(x_t, \pi(x_t))\right\}$$

Solving for optimal policy requires dealing with nonlinear equation

$$\pi^*(x) = \arg\min_{\pi} \left\{ R(x, \pi(x)) + E\left\{\sum_{t=0}^{\infty} \gamma^t r(x_t, \pi(x_t))\right\}\right\}$$

Solving for strategies is non-linear; we turn to solving for value (if G is terminal value)

$$V(x,t) = \min_{u \in U} \left[V(f(x), t+1) + G(x,u,t)\right]$$

by reasoning in terms of discrete time.

Alternately, in terms of discrete states y, as possible outcomes of state x, and introducing action a_t

$$V^* = \min_a \left\{ r(a,x) + \sum_y P(x,y)V(y) \right\}$$

where $\sum_y P(x,y) = 1 \quad \forall x \quad , \quad P(x,y) \geq 0$

We iterate on V since the problem is linear. Let ζ_t be the proxy for V at time t ; we iterate thus, using this discretization :

$$\zeta_{t+1}(x) \longleftarrow min\left(r(x,\pi) + \gamma.\sum_y P(x,y)\zeta_t(y) \right)$$

2.4. Enhancing realism with Qualitative modeling and Utility functions

The simple wealth effect reflected in instantaneous holding of wealth W_t is not enough. It is necessary to add an (psychology, risk averse) utility function U_2 that will "score" positively whenever risk prevention is done, regardless of immediate returns.

$$V_t = \max_\pi \mathsf{E}\left\{ \int_0^t e^{-rs} U_1\left[W_s^\pi\right] ds + U_2(\xi) \right\}$$

In real life, one can choose other types of objective functions (average past wealth, terminal wealth etc.). Determination of U_2 can be done through interviews of Top management using classic techniques of mapping risk aversion.

We arrive thus at a Hamilton-Jacobi-Bellman equation. However, it does not make much sense to precisely solve for the optimal amount of capital spending in hedging all those risks. Calibration in an environment where not only data is scarce, but furthermore structural conditions may be changing, is far too difficult and may yield large sensitivity to errors.

Such a model's goal is to highlight the interaction between various parameters, variables, constraints and allocation keys. In particular, it may highlight hedging long term, very rare risks (eg. earthquakes) whose return on investment cannot be immediately measured. Usually, accumulated discounted wealth is to be maximized.

2.5. Solution search: from HJB to Particle Filters

In theory, for solving this stochastic optimization problem to get the best policy leading to the best value $V^{\pi*}$ we can choose between:

- Solving as pure Hamilton-Jacobi-Bellman (Galerkin decomposition and viscosity solutions). Note that this decomposition reminds of Basel II matrix in AMA or Loan portfolios.
- Using neural techniques/supervised learning, if experience base available and sufficient.
- If the learning base is not sufficient, we can try Adaptive, Reinforcement Learning, Temporal Difference learning, or Q-learning. This last technique (see Watson) is particularly easy to set up and is model-free, at a cost of precision

$$Q_{t+1}(s,a) = g(s,a) + p(s,y).Q_t(y,a)$$

In Temporal Differences algorithms (see Tsitsiklis & Van Roy), the eligibility vector r_t serves as a memory of past states and corrections.

$$d_t = r(t) + \gamma V(t+1) - V(t) \qquad\qquad r_t \longleftarrow r_t + \gamma.d_t z_t$$

- Transpose real time filtering techniques :

- One of the easier ways in filtering is to try some moving average with or without weighting systems that let one attribute a decaying role to older data as one goes back in time.
- One of the more sophisticated filters is the Kalman filter. It covers the whole distribution of probability, however only if it can be assumed that probability is Gaussian, and that distortion/noise is linear. But in risk situations that are extreme, neither of these assumptions holds.

- Use Markov chain Monte Carlo and Sequential Monte Carlo exploratory techniques. Instead of backward fitting from data, this family of methods provide us with the flexibility of model-free techniques, the economy of calibration in a scarce data world, and with an ease of set-up. Sequential Monte Carlo (SMC), also called Interacting Particle Systems (IPS) filtering, has been applied to Probabilistic Robotics [Thrun][De Freitas] or Signal processing [Doucet]. A very great advantage is that scarceness of data can be remedied by learning from simulations, according to results stated specifically for Interacting Particle Systems (N. Shephard & Flury); but this is valid if and only if one can make sure the model has no bias.

3. TOY MODEL BASED ON MARKOV CHAIN AND MONTE CARLO SIMULATION

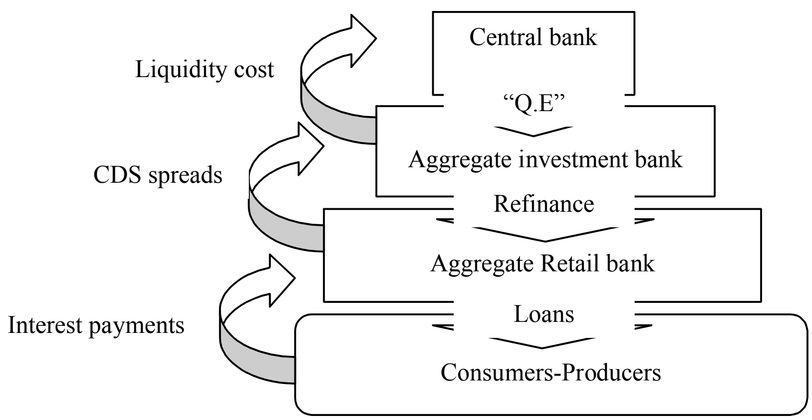
3.1 A simplified model of bank

Assets			Liabilities		
Reserve		β* K	Capital		K
Loan		L	Interbank debt		D
CDS sold			CDS bought		
« Cash »	Interest Received	ρ * L			
	Spread paid	ρ *(1+e) * L			
	« QE » or Newloan	Poisson(λ)			
	Discount of securitised EAD				

We use the following working simplified hypotheses:

- Loans are static except when securitised, then their value go to zero. But Total assets value is estimated by credit value adjustment (CVA)
- Whenever a Loan CVA in bank's portfolio goes below a threshold η (e.g. 0.85) , it will securitise away (buy a CDS), paying a premium based on current interest rate plus a spread. This spread being variable : in this simple model, it can vary as a function of total number of securitisation (CDS)
- When the Liquidity ratio goes below a threshold, corresponding to the Liquidity Coverage Ratio in Basel III, it will borrow from interbank market and pay an interest rate for this nominal principal.
- A CDS bought will allow a bank to discount and convert the Exposure at default (EAD) into Cash, thus easing its needs for liquidity (LCR)

3.2 Dynamics of the system



- The Economy is composed of a consumer-producer system, and a banking system.
- The banking system lends money to the consumer producer and gets paid with interests – determined by the "real sphere", a dependent economy in the neokeynesian framework.
- The banking system is composed of :
    - The aggregate retail bank which interfaces with the "real sphere" of producers-consumers , and who can buy default protection and liquidity fro :
    - The aggregate investment bank, which interfaces on the one hand, with the aggregate retail bank; and on the other hand, with :

- o The Central bank, who will inject liquidity by an optimal amount at each period. It is this optimal amount that we seek to determine.

The system is driven in time and at each turn of the clock; the following amounts are determined by random drawing:

- The CVA of each loans in the aggregate bank portfolio
- The generation of new loans, but depending on a Poisson process whose average λ grows inversely related to the interest rate ρ : the higher ρ , the less new loans are generated.
- Noise between the real liquidity needs of the aggregate investment bank and the perceived liquidity needs as observed from the central bank;

The following amounts are determined as subject to movements of the primary variables:

- The amount of interest that are paid to the retail banks by the “real sphere”; and the interests paid by to the investment banks by the retail banks as a result of their buying CDS protection.
- The second rate is higher than the first by a spread, which grows with the total number of securitizations.
- The amount of liquidity injected by the central bank as a function of its perception of liquidity needs.

3.3 Some early (to be confirmed) results

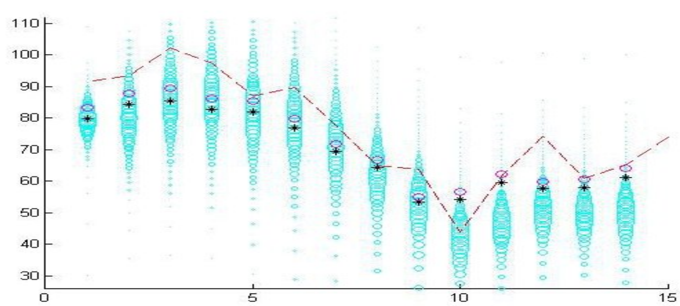


Fig. 2 : An example of particle systems tracking a random movement

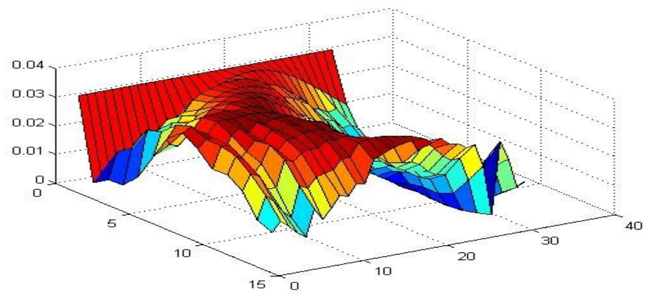


Fig. 3 Smooth case: gaussian distribution, noise in receiver, jitters in particle motion, smooth sine evolution, no jumps

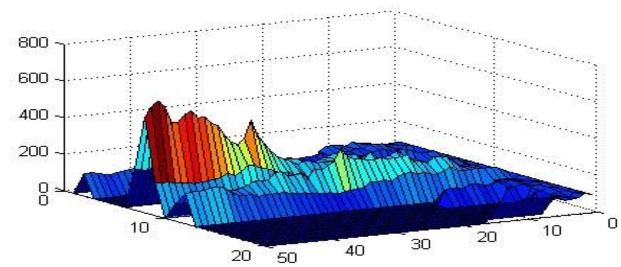


Fig. 4 Some of the 20 loans outperform, some others go to zero value (securitised) (Warning : time axis Right to Left)

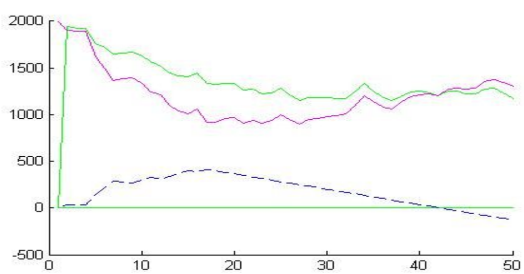


Fig 5 Evolution of Loans (green) total CVA (magenta) Cash (dash blue)

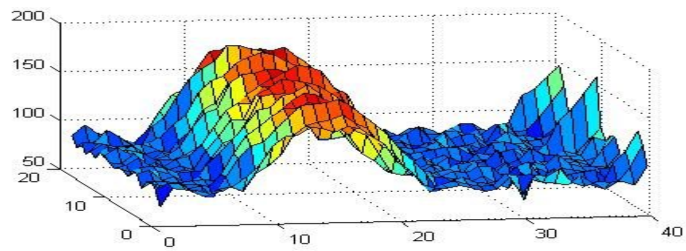


Fig 6 When cycle is high, no action. When in trough, lost amount is filled up by Q.E

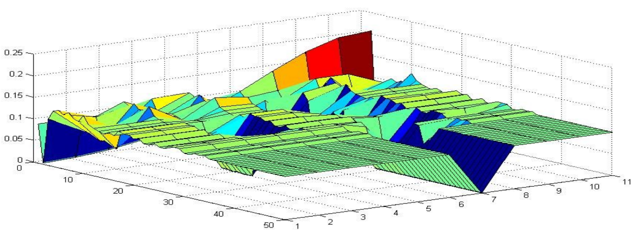


Fig 7 : Unconclusive shape of posterior distribution at this stage

4. IN SUMMARY

Most of Basel II and III financial regulations are based on data, treated in a statistical mindset. While regulators call for forward-looking risk management, they suggest no time-based modeling approach. In macroeconomic catastrophes modeling, this is a great setback. This paper reports the experimental application of combined forward looking methods, specifically, filtering techniques, like Sequential Monte Carlo, or Interactive Particles systems, with stress testing causal scenarios in systemic context.

Basel II framework and models leave out temporal dynamics in risk management processes in banks. We introduced stochastic equations to solves deficiencies of existing, risk non-sensitive, models. The new models are proposed as value-based, time varying, functions of rare catastrophic scenarios allowing arbitrage between risk mitigation decisions. Next, affiliation is established from Stochastic Optimal Control foundations, through Reinforcement Learning and Temporal Differences Learning, to this toy model set of equations.
We use this theoretical approach to introduce a simplified prediction process in a Bayesian framework, and to model economic forecast and rationalized control (e.g. by a Central bank). Some parts of this framework borrow from Markov and related Bayesian inference techniques that underlie Interactive particle systems filters.

As an illustration, we examined the impact of exogenous shocks on a toy model of a banking system in an exposed economy. The national banks, characterized by their stylized balance sheet reduced to a vector of ratios, have

interbank borrowing and lending relationships. This constitutes the mechanism for spreading risks. Propagation is simulated through markov chains of random processes.
We observe the consequences of stressed perturbations in 2 directions:

- •Credit default (counterparty risk)
- •Liquidity (lack of confidence and lending) risk

We showed a toy model evolved from a dynamic framework, using adaptive learning (by establishing a common mutual interbank fund at an optimal level as buffer capital).

Basel II and Basel III regulations on macroprudential systemic risks (liquidity and default) can benefit from this new, exploratory, rather than statistical, approach to Risk Capital. The Basel II and III frameworks prescribe extensive back test and stress test of models of credit and operational risks. A question that generally arises is how to exhaustively test the models, and cover the important contingencies. Without a general framework, like the previously shown HJB equations comprising rational and control variables, it is indeed very difficult to fit into a same "big picture" disparate issues such as shop floor fire prevention and quantification of internal control capacities and interbank liquidity management.

While this kind of objective function can be used to lay the groundworks for a search of a coherent corporate-wide policy of risk management, it is extremely complex to obtain all the parmeters. It is certainly not necessary to solve the HJB equation completely in its glorious complexity to do effective risk management, but reflecting on a simplified version surely help keep in mind all the numerous factors, parameters and variables to arbitrage choices in risk management decision, in a rational manner.

Now with the possibility to lean from simulations of unbiased Interactive Particle Systems models, we may have an important tool to model, measure and manage whole systems of interacting banking risks.

REFERENCES

Beard, R.W., G.N. Saridis and J.T. Wen (1995) Approximate Solutions to the Time-Invariant Hamilton-Jacobi-Bellman equation, Journal of Optimization theory and Applications,

Castillo, M.T, G. Parrocha, (2002) Stochastic Control theory for Optimal Investment.

El Ghanjaoui S., K.H. Hvistendahl (2002) A Markov chain approximation scheme for singular investment-consumption problem with Lévy driven stock prices..

Framstad, N.C., B. Oksendahl and A.Sulem (1999), Optimal consumption and portfolio in a jump diffusion market with proportional transaction costs, Rapport de recherche n° 3749, Projet MATHFI, INRIA Rocquencourt, Paris.

Hojgaard, B., M.Taksar (1998). Optimal proportional reinsurance policies for diffusion models, Scandinavian Actuarial Journal, 81,166-180.

Lamberton, D., B. Lapeyre (1997) *Introduction au Calcul Stochastique appliqué à la Finance*, chap.7, Ellipses, Paris

Moody, J. (2003) Risk, Reward & Reinforcement, AMS Workshop- Machine learning , Statistics & discovery, presentation slides, Utah.

Mnif, M., A. Sulem (2001). Optimal risk control under excess of loss reinsurance, Rapport de recherche n° 4317, Projet MATHFI, INRIA Rocquencourt, Paris.

Tsitsiklis, J., Van Roy, B., (1999) Average cost temporal difference learning, Elsevier, Pergamon press, Automatica

Barto R., Sutton A., (1998) Reinforced Learning, an introduction, MIT Press, Bradford books

Thrun, S., Burgard, W., Fox, D., (2005), Probabilistic robotics, MIT Press,

Doucet, A., Johansen, J., (2008) A tutorial on Particle filtering and smoothing

Shephard, N., Flury, T., (2009) Learning and filtering via simulation: smoothly jittered particle filters," Economics Series Working Papers 469, University of Oxford.