Market Risk and Volatility Weighted Historical Simulation After Basel III

Jean-Paul Laurent[1] & Hassan Omidi Firouzi[2]

This version 28 December 2017

Abstract

Regulatory capital requirements for market risk, also known as the Fundamental Review of the Trading Book (FRTB), were disclosed by the Basel Committee on January 2016. This major overhaul of the Basel 2.5 framework challenges risk model specification and backtesting. Given the prevalence of historical simulation approach within large financial institutions, we focus on the Filtered (Volatility Weighted) Historical Simulation (VWHS) approach associated with a EWMA volatility filter. Volatility dynamics is then directed by a single parameter. We discuss how this decay parameter, chosen within a reasonable range, at banks' discretion, impacts capital metrics, backtesting statistics, as prescribed by the Basel Committee, and fouls the regulatory benchmarking of internal risk models. We show a trade-off between the resilience of risk models to periods of turmoil and the magnitude of capital metrics. Under the new regulatory rules, this would favour plain historical simulation, as compared with filtered or volatility weighed historical simulation. Understanding why, might be helpful for regulated banks, regarding the management of their market risk models, and supervisors involved in internal model approval.

JEL Classification: G18, C51

Keywords: Basel III, Fundamental Review of the Trading Book, Market Risk, Historical Simulation, Backtesting, Capital Requirements.

1. Introduction

Modelling market risk is widely documented, both regarding VaR or Expected Shortfall (ES) estimation and backtesting methodologies. However, the set of rules disclosed by the Basel Committee in January 2016, entitled "*Minimum capital requirements for market risk*", colloquially known as FRTB or Fundamental Review of the Trading Book[3], leads to a reassessment of existing approaches. Besides, regulators have expressed concerns about excessive variability of risk weighted assets associated with market risk. Outcomes of risk models do depend upon modelling choices. This paper aims at unveiling implications of discretionary choices regarding risk models, capital market requirements, backtesting and risk comparability under the new regulatory framework. Our research focuses on the speed of adjustment of the risk measure to the arrival of news, i.e. the decay factors in filtered or volatility weighted historical simulation methods. These parameters are set at banks' discretion leading to a trade-off between resilience of risk models during turmoil and conservativeness of the capital metrics. The new Basel 3/FRTB framework is intended to strengthen existing regulatory requirements, with a shift from a blend of current VaR and stressed VaR to a full-fledged stressed Expected Shortfall.

---

[1] Jean-Paul Laurent, University Paris 1 Panthéon – Sorbonne, 17, rue de la Sorbonne, 75005 Paris, PRISM Sorbonne & Labex RéFi, Paris, France, laurent.jeanpaul@free.fr , jean-paul.laurent@univ-paris1.fr

[2] Hassan Omidi Firouzi, former postdoctoral fellow at Labex RéFi, Paris, France, homidifirouzi@escpeurope.eu , h.omidi.firouzi@gmail.com, Enterprise Model Risk Manager at Royal Bank of Canada. This work was achieved through the Laboratory of Excellence on Financial Regulation (Labex ReFi, ANR-10-LABX-0095). It benefited from a French government support managed by the National Research Agency (ANR) within the project Investissements d'Avenir Paris Nouveaux Mondes (investments for the future Paris-New Worlds) under the reference ANR-11-IDEX-0006-02. The authors thank David Murphy, Matthew Pritsker and the participants of seminars held at Paris 1 Panthéon – Sorbonne university and European Institute of Financial Regulation for useful feedback. The usual disclaimer applies.

[3] https://www.bis.org/bcbs/publ/d352.htm

However, these new features, the use of an endogenously determined stressed period and the backtesting features associated with the FRTB, might exacerbate the trade-off between resilience and the required amount of capital. While filtered historical simulation methods are often privileged by academics, new regulations might further boost the use of simple historical simulation. If filtered or volatility weighted historical simulations approaches were to be used, they should be implemented taking into account the specificities of the new regulatory framework.

The importance of our findings is related to the prevalence of historical simulation methods among most large financial institutions. Even under the simplest case (no volatility rescaling of shocks on risk factors), we need to deal with the length of the estimation period (somehow more constrained under the incoming regulatory framework), the structure of the shocks to account for heteroscedasticity in the risk factors and the granularity of risk factors. When considering volatility weighted historical simulation, one needs to rescale the past observations of risk factors to account for volatility dynamics during the estimation period. Plain historical simulation is a special case, where no rescaling is processed. Volatility weighted historical simulation is known to have good back-testing properties at the price of increased pro-cyclicality. It has also been considered in a CCP context, for initial margin models, but the regulatory prescriptions are quite different regarding market risks within banking institutions. We show that the decay factor involved in the EWMA volatility filter, routinely used by financial institutions, can be estimated in several ways with a large range of outcomes. We suggest that this parameter should be closely monitored. How this modelling choice impacts the backtesting performance, the resilience of the risk model under stress and the amount of regulatory capital in a Basel 3/FRTB context should be clearly understood. We think that this will be helpful in monitoring model risk in normal and stressed times, both for financial intermediaries and their supervisory bodies.

Quantitative risk methodologies for market risks used to be widely supported by regulators and financial institutions. The 1996 Basel amendment for market risk[4] and JP Morgan RiskMetrics[5] are cornerstones of this shift to risk models. However, the 2008 crisis showed major limitations, such as clustering of VaR errors and large overshoots of market losses above VaR[6]. The introduction of Stressed VaR and the use of multipliers was meant to deal with previous issue and with the underestimation of risk during periods of low volatility; the so-called "Minsky moments" recently investigated by Danielsson et al (2016). These fixes, part of the Basel 2.5 package, are still in force.

In January 2016, the Basel Committee disclosed a document setting up the new regulatory framework for the computation of risk weighed assets within the trading book[7]. This is the outcome of a process initiated in May 2012 (http://www.bis.org/publ/bcbs219.htm), still ongoing regarding key technical features[8].

[4] http://www.bis.org/publ/bcbs24.pdf

[5] http://yats.free.fr/papers/td4e.pdf

[6] In the summer of 2007, David Viniar, Goldman Sachs CFO stated "We were seeing things that were 25-standard deviation moves, several days in a row", https://www.ft.com/content/d2121cb6-49cb-11dc-9ffe-0000779fd2ac . This led to harsh criticism regarding the ability of risk models to cope with extreme market conditions (Dowd et al. (2008), Danielsson (2008), Crotty (2009), Dowd (2009)). To be fair, David Viniar did not mention whether he was comparing market moves to long-term historical standard deviation or current estimates of standard deviation, nor that Goldman Sachs assumed (conditionally) Gaussian returns when computing risk measures for its own purpose. However, it is quite likely that some banks experimented series of VaR exceptions in a row, with large overshoots during market turmoil.

[7] Minimum capital requirements for market risk, 2016, Basel Committee on Banking Supervision, http://www.bis.org/bcbs/publ/d352.pdf

[8] As part of its 2017-18 work program the Basel Committee planned some "Targeted adjustments and simplifications to the revised market risk and securitisation frameworks" and revived the former "Trading Book

Excessive variability of risk measures provided by internal bank models (advanced approaches in the US) has been under scrutiny: The Basel Committee conducted benchmarking exercises based on hypothetical portfolios as part of it RCAP (Regulatory Consistency Assessment Programme); large "unexplained" variability needs to be better understood[9]. Lack of trust in risk models jeopardizes comparability among banks, shatters confidence in bank disclosures regarding their risk exposures and thus casts doubts about the resilience of the banking system during market turmoil. The new regulatory framework factors in criticisms about excessive variability of risk weighted assets for market risk and suspicion about banks gaming the rules[10]. Updated requirements associated with the use of internal risk models have been considered:

- Eligibility to using internal models is only provided at desk-level.
- Data requirements have been enhanced, including stringent criteria for risk factors to be considered as "modellable" (i.e. valid inputs for internal risk models).
- The alignment between models used by the front office and for risk measurement is closely monitored through the so-called P&L attribution tests.
- Compulsory disclosure of risks computed under the standard approach is part of continuous monitoring of the outcomes of internal models for market risk.

Besides using a capital metrics based on a 97.5% Expected Shortfall, the FRTB departs from existing Basel 2.5 setup:

- The Expected Shortfall will be computed only over a stressed period, which is endogenously determined, and thus risk-model dependent[11]. The capital requirements for market risk are currently equally based on (current) VaR and stressed VaR[12].

Group" that drafted FRTB, now "Market Risk Group". On November 2016, the European Union disclosed its implementation of these new Basel rules https://ec.europa.eu/transparency/regdoc/rep/1/2016/EN/COM-2016-850-F1-EN-MAIN.PDF. Regarding backtesting issues, they closely mimic rules enacted by the Basel Committee. The US Treasury disclosed in June 2017 a report, https://www.treasury.gov/press-center/press-releases/Documents/A%20Financial%20System.pdf recommending delaying the implementation of the FRTB in the United States until it could be "be appropriately calibrated and assessed". So far, no advance notice of proposed rulemaking (ANPR) has been disclosed by the Fed. However, in December 2017, the US committed to implement the upcoming capital requirements for market risk whose implementation has been delayed from 2019 to January 2022: https://www.bis.org/press/p171207.htm

[9] http://www.bis.org/publ/bcbs240.pdf, http://www.bis.org/publ/bcbs267.htm https://www.eba.europa.eu/documents/10180/15947/20131217+Report+on+variability+of+Market+RWA.pdf

[10] Colliard (2014) mentions JP Morgan shift of IRC model associated with the London whale CDS trade inventory. See also the US Senate inquiry on that topic: https://www.hsgac.senate.gov/subcommittees/investigations/hearings/chase-whale-trades-a-case-history-of-derivatives-risks-and-abuses

[11] This will the case if the reduced set of risk factors equals the full set of risk factors. Otherwise, the ratio of current expected shortfall, computed over the full set of risk factors, to current expected shortfall computed over the reduced set of risk factors is involved, which leads to some dependence with respect to current market conditions.

[12] Basel Committee (http://www.bis.org/publ/bcbs158.pdf) mentions that stressed period should be "*relevant to the banks' portfolio*". But in many cases, the 12 Month stress period includes September and October 2008. The FRTB is more prescriptive than Basel 2.5: "*The expected shortfall for the portfolio using this set of risk factors, calibrated to the most severe 12-month period of stress available over the observation horizon, is calculated*". Detailed requirements will likely depend upon national jurisdictions and discretionary model supervision.

- Regulatory backtesting will be conducted at a more granular (desk) level and involves a set of indicators, daily VaR exceptions at both 97.5% and 99% levels, p-values, daily 97.5% one day 97.5% Expected Shortfall[13].
- The capital metrics involves computation based on shocks on risk factors over a base liquidity horizon of 10 days (FRTB §181 c). No square root scaling-up from the daily backtesting horizon to the 10 days capital metrics horizon is permitted.

Though the FRTB is rather prescriptive regarding the use of data (one-year stressed period for Expected Shortfall computations) and backtesting (past year), there are some discretionary (internal) modelling choices: risk factor granularity, parametric vs historical models, features of shocks on risk factors, volatility rescaling, modelling of correlation. The number of market risk factors could range from around 5 000, roughly the number of risk factors involved in the standard approach, more often in between 50 000 and 100 000[14]. For ease of exposition, we will thereafter focus on a long S&P500 equity portfolio. Besides its intrinsic importance and wide use for benchmarking risk models, we intend to investigate key modelling choices and their implications regarding backtesting and capital metrics, that are unrelated to portfolio complexity.

We do not aim at determining the most suitable data generating process. We rather assess the implications of actual modelling choices made by banks on the measurement and assessment of risk models and on model risk. Most of these approaches, such as parametric GARCH, including GJR-GARCH(1,1) used for initial margin models by a number of CCPs, GARCH-EVT, RiskMetrics/EWMA, historical simulation and its variants (Filtered Volatility Estimation, Volatility Weighted Historical Simulation or VWHS) are location-scale models: the data generating process involves a conditional mean, a conditional standard deviation and some distribution for the innovations. Given the usually short-term horizon involved by regulation, 10 days for market risk, one can reasonably consider that conditional mean has a marginal impact on the computation of risk measures[15].

Given the degree-one positive homogeneity of VaR or Expected Shortfall, the risk measure can then be written as the pointwise estimate of volatility times the risk measure of innovations. Innovations could follow some given iid distribution, such as Student t, laxer assumptions allowing the use of EVT might be made. Non-parametric approaches to the risk measure of innovations are also being used, as in the Volatility Weighted Historical Simulation approach introduced by Boudoukh et al (1998). In all cases, the derivation depends on the chosen estimation technique and perhaps more importantly on the length of the data sample.

[13] Failure to comply with new backtesting requirements leads to ineligibility of trading desks to the Internal Models Approach (IMA) and a fall-back to the Standardised Approach (SA), associated with more stringent capital requirements. According to ISDA, based on the outcome of the June 2015 Basel Committee Quantitative Impact Study, the ratio of SA capital requirements to the risk measure derived from internal expected shortfall models would range from 2.1 to 4.6 depending on the risk factor class Industry FRTB QIS Analysis, https://www2.isda.org/attachment/Nzk0OA==/Industry%20FRTB%20QIS%20Analysis%20Executive%20Summary%20OCT%202015.pdf

[14] Under the Basel Committee prescriptions, internal models should be at least as granular as the standard approach, thus the number of risk factors associated with IMA is floored by the number of risk factors involved in SA. The number of risk factors currently involved in internal models (advanced approach) is not disclosed, thus the provided figures are indicative and could vary significantly from one bank to another. Since the non modellable risk factor (NMRF) charge might be large, there is a possibility that some banks would decide to reduce the granularity of risk factors when moving to FRTB.

[15] 10 days is the horizon associated with the most liquid market risk factors. Less liquid market risk factors have liquidity horizons up to 120 days, but this results from computations over 10 days, which are scaled-up.

Computing short-term prediction of volatility is thus of first importance. This usually involves a GARCH or I-GARCH (EWMA) dynamics. The January 2016 Basel document is not prescriptive about specification and estimation of volatility dynamics, for instance the length of the estimation period. This is left at banks' discretion. Picking-up a number, such as the .94 decay factor in EWMA RiskMetrics is also regulatory compliant. This discretionary approach to setting key parameters involved in risk models is not innocuous as will be further discussed.

Even though many academic articles focus on parametric or semi-parametric GARCH models, the approach predominantly used by main financial institutions is non-parametric, more precisely based on empirical quantiles of innovations, so called historical simulation. Due to increased computational constraints associated with the new Basel rules, it is likely that some banks still using parametric approaches (so called "Monte Carlo") will have switched to historical simulation by implementation date of the new Basel rules. As for volatility filtering, whenever being used, EWMA seems also predominantly being used by banks, in most cases with a constant over time decay factor of .94 applied to all risk factors, as prescribed in RiskMetrics document. Operational constraints such as a stable and timely estimation of GARCH models, involving 50 000 risk factors should not be misjudged. The good news of setting-up harsh constraints in volatility dynamics, such as using the same decay factor for all risk factors, is that the number of key and easy parameters to grasp and understand is dramatically reduced. Apart from a uniform treatment of all risk factors, the drawback of such a parsimonious and somehow arbitrary approach is that the decay factor becomes a parameter of first importance, regarding capital metrics and the resilience of the bank's risk model.

We discuss different estimation methods of the decay factor and show that this leads to large discrepancies both regarding estimated values and pointwise volatility estimates. The resulting ratios of pointwise daily estimates of volatilities typically range from .5 to 2 even though median and average remain close to one. On top of this, small autocorrelations lead to highly unstable volatility predictions. This translates to risk measures. Besides, considering quantiles of innovations casts doubts on the ability of parametric GARCH or GARCH-EVT to capture tail dynamics (see Harvey and Siddique (1999), Brooks et al (2005) for modelling of higher order conditional moments) and the validity of the stationary/ergodic assumptions required under EVT approaches.

We proceed with the benchmarking tools prescribed by the Basel Committee, number of VaR exceptions at 97.5% and 99% over past year and p-values. We also consider some conditional coverage checks. For a large range of decay factors, risk models do not fail back-tests, not a surprise given the well-known lack of power of usual statistical tests. Using smaller values of decay factors results in a small improvement during stressed periods, since risk estimates react more promptly to volatility surges. Unlike CCP initial margin models this does not lead to procyclical effects since the risk-measure is based on a stressed period, is averaged and VaR exceptions are only computed over the current period (past year). Using a volatility floor as in Murphy et al (2014) or Gurrola-Perez and Murphy (2015) would reduce the occurrence of VaR exceptions during so called "Minsky moments", i.e. calm before the storm. Unlike CCP initial margin models, this is not usually done at the expense of inflating capital requirements during low volatility periods, since capital metrics are computed under a stressed period, usually a volatile one.

The paper is organised as follows. Section 2 deals with the modelling and pointwise estimation of volatility. The focus is put on the intricacies of the estimation of decay factor in EWMA approaches. Section 3 documents the substantial impact of previous modelling choices on VaR estimates under the (volatility weighted) historical approach. Section 4 provides some empirical evidence regarding backtesting performance along the prescribed lines of Basel 3. We also consider econometric approaches associated with conditional coverage restrictions. Due to lack of power of statistical tests, a wide range of decay factors and thus internal modelling choices, will not be rejected. However, focusing on stressed periods, it appears that plain historical simulation underperforms VWHS, with clustering of VaR exceptions. More generally, the use of volatility filters that react promptly to large market moves, i.e. smaller values of decay factors, is associated with more resilient market risk models. Section 5 deals with the computation of capital metrics, i.e. 10 days 97.5% Expected Shortfall computed over an endogenously stressed period. We suggest an implementation of VWHS that conforms with regulatory prescriptions and with the Basel III backtesting metrics. We discuss the impact of the decay factor on the determination of the one-year stressed period and the associated expected shortfall. The endogenous determination of the stress period, overlapping 10 days returns and the use of a non-robust risk measure challenge the usefulness of VWHS, as compared with plain historical simulation. We conclude with comments regarding the development and use of internal market risk models under the January 2016 rules disclosed by the Basel Committee.

2. Dynamics of Volatility

We will thereafter focus on the S&P500 index confused with the studied risk factor. In the univariate case, the data generating process for the risk factor often takes the form of a location scale model, i.e. $r_t = \mu_t + \sigma_t z_t$. r_t will thereafter be a relative return, though log-returns might also be considered. In the case of a constant elasticity of variance diffusion process, a rescaling of prices would be required to cope with heteroscedasticity. Proper definition of "shocks" on risk factors is an important issue in risk modelling, but will be left aside in this paper.

Given the short time horizons, typically one day for backtesting under the Basel Committee framework and 10 days regarding the capital metrics, we will moreover assume subsequently that the conditional mean $\mu_t = 0$. σ_t is the conditional (on available information at $t-1$) standard deviation and z_t relates to the innovation term.

Let us confuse for a while the horizon of the capital metrics and the periodicity of the returns, say one day. Given that standard risk measures considered by regulators, including VaR and Expected Shortfall are degree one positively homogeneous, we have $\rho(r_t) = \sigma_t \times \rho(z_t)$, where ρ stands for the considered risk measure[16]. Pointwise estimations of conditional volatility σ_t are thus of utmost importance. A benchmark model is GARCH(1,1), which mitigates parameter parsimony and a number of realistic features regarding clustering of squared returns:

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{1}$$

When $\alpha_0 > 0$ and $\alpha_1 + \beta_1 < 1$, the volatility process is stationary, thus broadly speaking there is some mean reversion towards a constant value. It is usually assumed that z_t are i.i.d. Then, all dynamics is captured by moments of order two. Different estimation strategies depend upon the specification of the distribution of z_t. Let us remark that under the previous framework the risk measure turns out to be a constant multiple of σ_t.

The GARCH(1,1) specification can be improved in many ways, for instance introducing asymmetric responses depending on upward or downward moves as in Glosten et al. (1993): Volatility tends to increase faster after large negative shocks as compared with positive shocks of same magnitude. Abad

[16] We chose to consider relative returns, thus $r_t = (I_t - I_{t-1})/I_{t-1}$ on a trading day where no dividends are being paid. Assuming that the S&P500 index is used to compute the daily $t-1$ risk measure on a S&P500 portfolio and that $1/I_{t-1}$ units of the index are being held, the change in value between the current date $t-1$ and the one-day horizon is provided by $(I_t - I_{t-1})/I_{t-1}$: r_t can actually be related to a change in value. In the case of one a 99% one-day VaR computed at date $t-1$, we would consider the 1% quantile of r_t.

and Benito (2013), Abad et al (2014) provide a review of potential extensions to the base model, usually at the expense of increasing the number of parameters.

In the following table, Table 1, we provide the parameters' estimates for GARCH(1,1) using historical data for S&P 500 index returns from 2/1/2002 to 30/12/2015 and assuming Gaussian innovations (rugarch package in R). The mean reversion parameter in GARCH models, α_0 is close to zero.

Parameters	Estimate	Std. Error	t value	$Pr(>\|t\|)$
α_0	0.000002	0.000001	1.8454	0.064981*
α_1	0.090674	0.012132	7.4739	0.000000
β_1	0.895003	0.012817	69.8308	0.000000

Table 1: Data used from 3/1/2002 to 31/12/2015 to estimate GARCH(1,1) parameters.

Given low values of α_0, it is not a full surprise that, unlike in the financial econometrics literature, the financial industry has privileged a more parsimonious EWMA filter for computing pointwise estimates of volatility:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2 \tag{2}$$

The dynamics of volatility depends on a single parameter, λ, the decay factor. The previous dynamics is also known as I-GARCH(1,1). In contrast to the previous modelling, there is no mean-reversion, the dynamics is not stationary. Oanea and Anghelache (2015) review the use of EWMA approaches: Tse (1991) and Kuen and Hoong (1992) emphasize the fact that EWMA model overperformed ARCH models in estimating the risk for Japanese and Singaporean financial markets. On the other hand, Hammoudeh et al. (2011) found that GARCH-t model overperforms EWMA in estimating the risk involved in commodities' markets. Similarly, Degiannakis et al. (2011) show that the ARCH framework is better in estimating risk compared to RiskMetrics model. Chen et al (2012) report also a poor forecasting ability of EWMA approach in a Value at Risk context. Guermat and Harris (2002) propose a robust EWMA approach involving a different recursive scheme.

One can also notice that there is no strictly positive lower bound on volatility, which might be an undesirable feature. Long periods of low squared returns result in very low estimates of current volatility. For this reason, some CCPs floor the estimated volatility when computing initial margins; this also helps dealing with pro-cyclicality issue, further discussed within this paper, in the Basel III/FRTB context.

We can expand equation (2) as:

$$\sigma_t^2 = \lambda^T\sigma_{t-T}^2 + (1-\lambda)\sum_{i=1}^{T}\lambda^{i-1}r_{t-i}^2 \tag{3}$$

The conditional variance involves a weighted average of past squared returns, a computation period T and a conditional variance seed σ_{t-T}^2. There are no clear-cut choices. For large T and λ far-away from 1, the value of the seed will not have great importance, but we may be faced with some practical issues as further discussed.

One key issue to be considered is the determination of the decay factor. Under the RiskMetrics' approach (J.P. Morgan (1996)), the decay factor was set to $\lambda = 0.94$ for daily returns (further on conditional returns were assumed to be Gaussian). The estimation of decay factor was based on the root mean-squared prediction error (RMSE), i.e.:

$$\hat{\lambda} = \underset{\lambda \in (0,1)}{\arg\min} \frac{1}{T} \sum_{i=1}^{T} \left[\sigma_i^2(\lambda) - r_i^2 \right]^2 ,$$

where $\sigma_i^2(\lambda)$ is set according to equation (3) and T is the length of the estimation period. While the RiskMetrics technical document reports a wide dispersion of estimated decay factors for different asset classes, it was decided to retain a single number of 0.94 whatever the considered period or the considered risk factor. Issues related to EWMA RiskMetrics approach have been outlined by Fan and Gu (2003) and González-Rivera et al (2007).

Alternative criteria have been considered regarding the decay factor's estimation. Fan and Gu (2003) use a pseudo-likelihood method: The decay factor is computed as:

$$\hat{\lambda} = \underset{\lambda \in (0,1)}{\arg\min} \frac{1}{T} \sum_{i=1}^{T} \left(\log\left(\sigma_i^2(\lambda)\right) + \frac{r_i^2}{\sigma_i^2(\lambda)} \right) .$$

González-Rivera et al (2007) estimate decay factor(s) by minimizing the check loss function (scoring function) as follows:

$$\hat{\lambda} = \underset{\lambda \in (0,1)}{\arg\min} \frac{1}{T} \sum_{i=1}^{T} \rho_\alpha(e_i),$$

Where $e_i = r_i - q_\alpha$ and $\rho_\alpha(e_i) = \left(\alpha - 1_{\{e_i<0\}}\right)e_i$ for $\alpha \in (0,1)$. q_α is the quantile at confidence level α. González-Rivera et al (2007) find that when considering a 99% 10 day - VaR and equity markets, the RiskMetrics approach tend to overestimate the decay factor(s).

Estimation method/ length of historical data	10 years	First 5 years	Second 5 years
Squared error method	0.8992854	0.8207192	0.9030331
Pseudo likelihood method	0.9331466	0.9525935	0.9146936
Check loss method at 1% level	0.9010942	0.9406649	0.8398029
Check loss method at 2.5% level	0.8829908	0.9557358	0.8634209

Table2: Decay factor estimation using three different methods and different periods of historical data. First five years: 1/2/2003 – 12/21/2007. Second five years: 1/2/2008 – 12/31/2012. Ten years: 1/2/2003 – 12/31/2012.

Table 2 reports decay factor estimates using the above three different approaches (and two confidence levels, corresponding to the Basel 3 backtesting requirements for the check loss function) and different estimation periods.

The most obvious outcome is that there is a wide range of plausible values for the decay factor, depending both on the chosen criteria and the estimation period. Moreover, consistently with the previous statement, a closer inspection of the (one-dimensional) criteria shows that they are essentially flat.

Figure 1 compares one day ahead volatility predictions based on decay factors equal to 0.94, recommended in RiskMetrics and mentioned as a benchmark in many textbooks, and 0.8[17]. The y-axis reports the ratio of volatility estimates over the ten years 2003 – 2012 period. The median is not far

[17] The volatility estimates have been computed with one year rolling period and a seed equal to the first squared returns. The daily update and the choice of a one-year period are inspired by FRTB. Using a more conventional approach, i.e. leaving the EWMA update operate throughout the ten years period would not have a material impact on our results. Given the 0.8 and 0.94 values of the decay factor, the seed one year away from the volatility estimation date has a quite small impact.

away from one. This is not full surprise since EWMA estimators involve an averaging of past squared returns, with weights summing-up to one. However, the interdecile range shows a rather large dispersion of the ratios and a strong dependence on the considered date. Lopez (1999), Christoffersen and Diebold, (2000), Angelidis et al (2007), Gurrola-Perez and Murphy (2015) also pointed out the issues related to volatility prediction in a risk management context. Building volatility filters is even more intricate when considering different risk factors (Davé and Stahl (1998)).

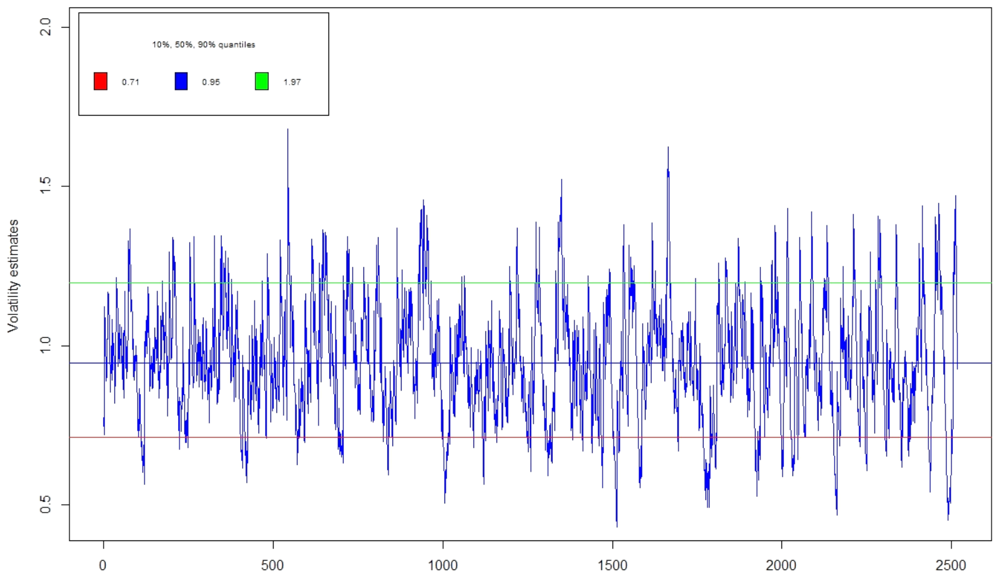

Figure 1: Ratio of EWMA one day ahead estimates of volatility with decay factors equal to .8 and .94 over the ten years 2003 – 2012 period. Green line corresponds to the 90% quantile, red line to the 10% quantile.

3. Implementation of Risk Measures

3.1 Historical Simulation and Banks' Market Risk Internal Models

Most of the academic literature focuses on parametric implementation of risk measures. The use of GARCH dynamics is prominent and a wide range of parametric distributions regarding innovations are considered. The semi-parametric GARCH-EVT is also well documented. Though CCP do not usually publicly disclose details of their initial margin models, anecdotal evidence does not discard GARCH models (Murphy et al (2014)).

A decreasing number of banks are reported to use "Monte Carlo" approaches. We think that this denomination is misleading, since it usually involves a parametric model for risk factors, a multivariate normal distribution for instance. Monte Carlo techniques are only required when it comes to computing the distribution of portfolio returns, especially when the valuation function is non-linear with respect to risk factors. In some cases, even for linear portfolios, the distribution of returns cannot be achieved analytically and requires simulation of risk factors.

Many large banks are reported to use non-parametric approaches, often called "historical simulation" (see Perignon and Smith (2010), Mehta et al (2012) and the study conducted by the European Banking Authority (2017)). O'Brien and Szerszen (2017) notice that on a small sample of five large US banks, three of them were computing VaR through historical simulation. Plain (i.e. non-filtered or volatility weighted) historical simulation involves the empirical distribution of past portfolio returns, typically one or two years in the current Basel 2.5 setting. Basel III computational requirements regarding capital metrics are more stringent, due to the endogenous computation of the stress period and increased granularity of regulatory metrics. Moreover, computing ES through a Monte Carlo approach with a given confidence level is comparatively most costly than computing a VaR, due to the robustness of the VaR, as compared with ES. This is likely to increase the predominance of historical simulation.

The dismissal of GARCH models is likely to be related to the huge number of risk factors, involved in banks' internal models, a typical number would be 50,000. Specifying, estimating, implementing and monitoring a stable and meaningful GARCH model is therefore a huge challenge. Due to the large number of parameters involved, constraining the model, such as the use of a small number of common factors is to be considered, which is not an easy task either.

One of the issues when using plain historical simulation is the choice of the length period, more often one or two years. Therefore, these models are subject to "ghost effects" when dates associated with the largest losses get out the sample. To mitigate this effect, Boudoukh et al (1998) proposed to assign more probability weights to recent events than those happened in the far past. This method has been known as BRW method in the literature. It improves over plain historical simulation though Pristker (2006) reports some drawbacks. Unlike filtered or volatility weighed historical simulation, the BRW method is not compliant with FRTB prescriptions: When, it comes to ES computations, the Basel Committee states that "*observations (…) must be equally weighted*".

3.2 Filtered and Volatility Weighted Historical Simulation Models

Another concern with HS is that these models poorly deal with the dynamics of volatility of returns (Engle and Manganelli (2004), Pritsker (2006), Andersen et al (2007)). Volatility of returns for a given risk factor can vary through time and considering it as a fixed constant through time might understate the estimated VaR. This feature can partially be accommodated by considering shorter time intervals at the price of increased volatility of VaR estimates, which is both problematic to regulators and practitioners[18].

Filtered historical simulation (FHS) has been introduced by Barone-Adesi et al (1999). An almost identical technique known as Volatility Weighted Historical Simulation (VWHS in the sequel) has been considered by Hull and White (1998).

To get a better flavour of FHS, consider the following time series modelling historical returns: $r_t = \sigma_t z_t$, where z_t is the series of i.i.d innovations. When σ_t is fitted with a volatility tracking model (say using a EWMA or a GARCH filter), we normalize the historical returns to get a set of r_t/σ_t. To estimate quantile for the next day, we bootstrap the standardized innovations and plug the simulated values into $r_t = \sigma_t z_t$. The VaR is then computed as the opposite of the quantile of this distribution.

Different improvements have been proposed regarding the statistical bootstrapping step. We refer to Mancini and Trojani (2011), for a description of a robust bootstrap method. Regarding VaR estimation, FHS has been supported by many articles in the literature. We can refer to Kuester et al (2006), Pritsker (2006), Marimoutou et al (2009), Zikovic and Aktan (2009). For instance, Pritsker (2006) compared the performance of plain historical simulation and BRW estimators with FHS and concluded that FHS is more accurate than both plain historical simulation and BRW. Kuester et al (2006) show that the FHS

[18] The March 2017 European Banking Authority benchmarking study, based on EU banks with approved internal models reports that 58% of respondents were using a one-year window to compute VaR, while other respondents were using lengthier windows.
https://www.eba.europa.eu/documents/10180/15947/EBA+Report+results+from+the+2016+market+risk+benchmarking+exercise+-+March+2017.pdf
FRTB favours using a one-year period, which can be shortened during times of stress, but cannot be less than six months.

model explained above along with its variants, like using the Skewed t Student distribution for simulations and at filtering stage, are the most robust to the choice of data window length.

VWHS also accounts for volatility changes: past returns are scaled according to a revolatile and devolatile process (Gurrola-Perez and Murphy (2015)), i.e. dividing past returns by the estimate of the volatility for the corresponding date and multiplying past returns by current volatility:

$$\tilde{r}_t = \frac{\sigma_T}{\sigma_t} r_t \, , 1 \leq t \leq T$$

Where σ_T is the most recent forecast of the volatility. Then the VaR is computed from the empirical distribution of $\tilde{r}_t$. Unlike FHS, there is no bootstrapping, but this feature apart FHS and VWHS coincide. O'Brien and Szerszen (2017) consider VWHS with a GARCH filter, Gurrola-Perez and Murphy (2015), Murphy et al (2014) discuss possible extensions of VWHS, including the use of a floor on volatility when using EWMA filter.

While VWHS principles are simple, we must keep in mind the following ingredients: the value of the decay factor, the data window length and the EWMA seed, i.e. the first value of the volatility in order to start the iterative process of computing volatilities. A "long-term average" of volatilities is often chosen, one could also start with the first squared return over the computational window. We will discuss other choices further on.

	EWMA volatility filter	GARCH volatility filter
Parametric innovations	RiskMetrics	Academic Standard
Non-parametric innovations	FHS / VWHS	Garch-EVT

Table 3: Approaches to VaR and ES computations.

The above table summarizes the different approaches to VaR or Expected Shortfall computations. We remind that, thanks to degree one positive homogeneity, we have in all cases:

$$\rho(r_T) = \sigma_T \times \rho(z_T)$$

In a standard parametric GARCH model, σ_T is estimated accordingly, the z_T are assumed to be iid, with a given parametric distribution, Gaussian, Student t or more involved ones such as reciprocal inverse Gaussian. The parametric GARCH model This can be considered as an academic benchmark and is also being used by some clearing houses for the computation of initial margins. RiskMetrics assumes that the z_T are Gaussian, but uses a EWMA filter to compute current values of the volatility σ_T.

As compared with parametric GARCH or to a lesser extent to GARCH-EVT models, VWHS does not make any formal assumption about innovations. As for other historical simulation techniques, the estimation window is of key importance, but the January 2016 Basel rules (§ 181(e)) clearly push towards using a one-year period, even though this might be reduced during periods of market stress.

We might have also considered other estimates of volatility, based on options' implied volatilities, such as the VIX index. The idea seems appealing due to the forward-looking nature of VIX, while still being "historical data" and amenable to internal modelling. Unfortunately, we further show that a rescaling of returns based on VIX may lead to poor backtesting results.

Figure 2 provides ratios of 99% 1 Day VaR estimates computed under VWHS with decay factors 0.8 and 0.94. over the ten years 2003 – 2012 period. For a given day T, we have considered the empirical quantile of the standardised returns $z_t = r_t/\sigma_t$ over the past year and multiplied it by the pointwise estimate of σ_T. The pattern mimics the one associated with pointwise volatility estimates. This confirms that the prominent role of volatility estimates in deriving market risk measures.

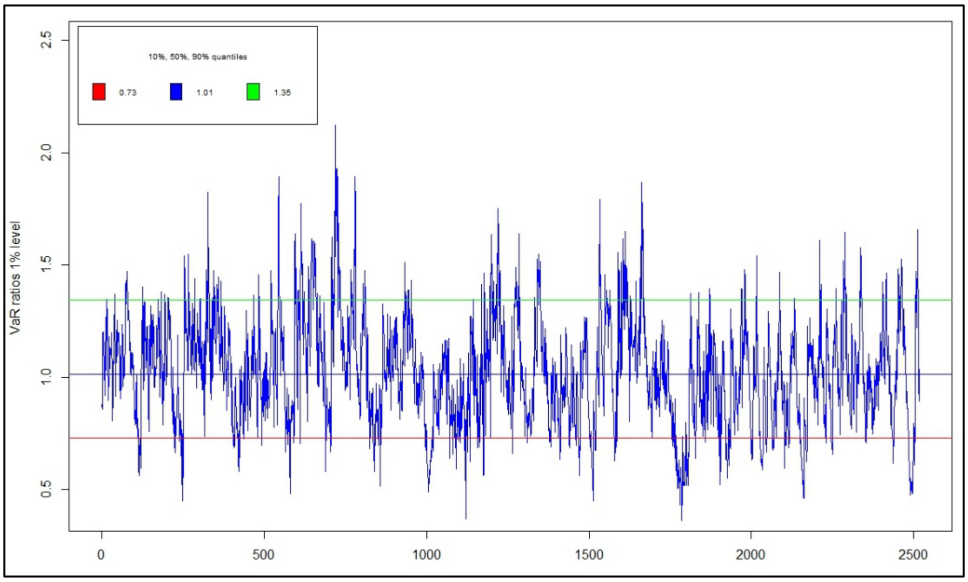

Figure 2: VaR1%/VaR1% for decay factors 0.8 and 0.94 respectively, trading days on the x-axis from January 2003 to December 2012.

The median of the ratios of VaR is almost equal to one, thus on average no modelling choice is more conservative regarding risk measurement. However, the ratio of the ninth to first deciles equals 1.85. Besides, this ratio is highly volatile. In other words, a modelling choice might look as aggressive on a given day and the converse will occur within a few days. Changes in risk models require supervisory approval and are thus sticky. Therefore, gaming capital requirements by discretionary updating of decay factor during turmoil would likely be deemed non-compliant with current regulations and the supervisory process[19].

The inherent large instability of VaR outcomes under some plausible modelling choices jeopardizes the benchmarking approaches based on hypothetical portfolios led by the Basel Committee and the European Banking Authority. Due to the “single day effect” it is hazardous to conclude to excessive variability of internal models’ outcomes and to disentangle what is due to “statistical noise” and what would come from deliberate choices aimed at relaxing regulatory requirements. In its latest benchmarking exercise, the European Banking Authority initiated an averaging over two weeks. This is a welcome move regarding the assessment of the variability of market risks weighted assets derived from internal models. A further enhancement would be using a 60 days averaging, consistently with the Basel Committee capital metrics.

The other component of VWHS VaR is the empirical quantile of the rescaled return. We did find some dynamics in the standardised returns r_t/σ_t as illustrated in Figure 3. We computed the ratio (VaR1%/VaR2.5%)/ (Φ^{-1}(99%)/ Φ^{-1} (97.5%) from January 2003 to December 2012. Φ stands for the Gaussian cdf. The VaR has been computed using VWHS and a decay factor of 0.8. The fluctuations in the ratio of VaR to inverse Gaussian CDF ratios away from 1 demonstrates that under VWHS the rescaled returns do not represent a stationary Gaussian distribution.

[19] As shown before, the estimates of decay factors are highly unstable. One could think of a dynamic decay factor model, as to lower decay factor during volatile periods. This goes beyond the scope of this paper.

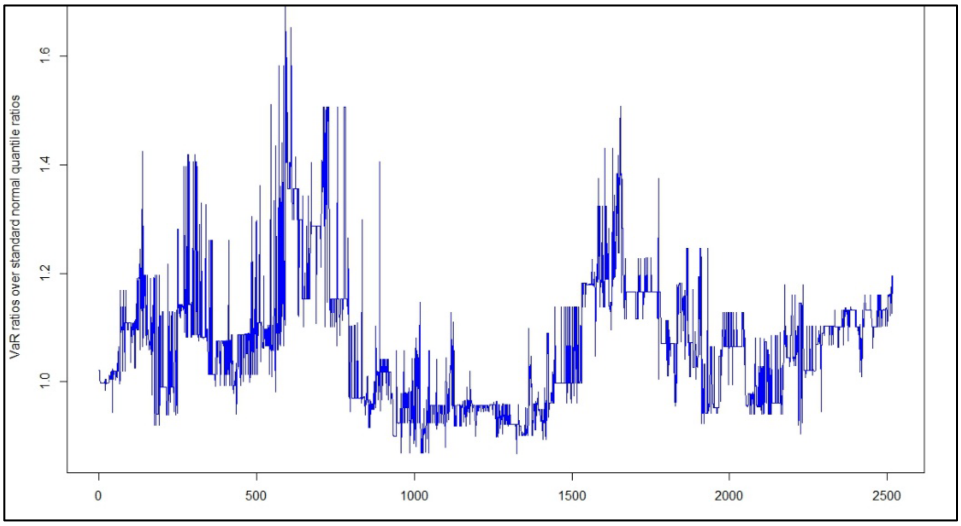

Figure 3: ratio (Var1%/VaR2.5%)/ (Φ^{-1}(99%)/Φ^{-1}(97.5%) on the y-axis, trading days on the x-axis from January 2003 to December 2012.

4. Impact Analysis of Choice of Decay Factor on Backtesting Metrics

The Basel Committee introduced some new backtesting requirements, based on the number of one-day VaR exceptions at desk-level and at 99% and 97.5% confidence levels. We show that a lower value of decay factor actually lowers the number of VaR exceptions under stress, thus being associated with better resilience. This is captured to some extent by conditional coverage tests. However, the lack of power of statistical tests provides some discretion in setting the value of decay factor, while meeting the regulatory backtesting requirements.

Backtesting VaR models is a standard issue. For a review of backtesting methods we refer to Berkowitz et al. (2008), Boudoukh et al (1998), Haas (2001), Hendricks (1996) and Kuester et al (2006). There is also a vast literature dedicated to the assessment of model risk due to parameter uncertainty or the choice of estimation method; see for instance Christoffersen & Gonçalves (2005), Alexander & Sarabia (2012), Escanciano & Olmo (2012), Escanciano & Pei (2012), Gourieroux & Zakoïan (2013), Boucher & Maillet (2013), Boucher et al. (2014), Danielsson & Zhou (2015), Francq & Zakoïan (2015), Danielsson, et al. (2016). We also remind that our focus is to concentrate on a key parameter left in the shadow, i.e. decay factor, and deal with the implications of different choices for risk management under Basel III/FRTB.

Let us denote by $I_{t+1} = 1_{\{r_{t+1} < -VaR_t(\alpha)\}}$ the binary variable associated with a "hit" or a VaR exception. The standard definition of VaR is equivalent to $E[I_{t+1}|F_t] = \alpha$. This conditional moment restriction is referred to as the conditional coverage property. The information set may content any relevant instrumental variables, such as past VaR exceptions. An implication of the conditional coverage property is $E[I_{t+1}] = \alpha$, so-called unconditional coverage (UC) property.

4.1 Backtesting Risk Models under the Basel 3 Framework

A test was set up to verify the UC property by Kupiec (1995) using the maximum likelihood ratio test under the assumption that VaR violations are independent. Lopez (1999) has shown that Kupiec's test and other alternatives have low power meaning that the probability of not rejecting an inaccurate VaR model is high.

In the 1996 amendment, the Basel Committee proposed to use a backtesting method known as "traffic light approach" to monitor VaR models (BCBS, 1996). As an implication of the unconditional coverage property and the independence between VaR exceptions, under the null hypothesis that the risk model is well specified, the number of VaR violations follows a Binomial distribution. The Committee then proposed to categorize the backtesting results for VaR at 1% significance level over the period of 250 days into three different zones as follows.

- Green zone, if the number of VaR exceptions is less or equal to 4.
- Yellow zone, if the number of VaR exceptions is between 5 and 9 inclusively.
- Red zone, if the number of VaR exceptions is larger or equal than 10.

The Basel III framework adds extra requirements. Firstly, backtesting metrics are requested to be reported at desk level (see Danciulescu (2010), Wied et al (2015) for a multivariate approach to backtesting) and not only at bank's portfolio level. The purpose is to validate internal risk models at a more granular sub-portfolio level. Desks that are not eligible to the internal model approach will be required to apply a standard approach, that is inspired by parametric approaches with conservative risk weights and correlation assumptions. More formally, each trading desk requiring internal model approval will need to produce the number of daily VaR exceptions over the past trading year at 99% and 97.5% confidence levels. The maximum permitted number of exceptions is equal to 12 at 1%, and 30 at 2.5%.

Columns 2 and 3 of Table 4 report the number of VaR exceptions over a period of 10 years (2005 – 2014) using different volatility filters. The expected number of VaR exceptions is equal to 25 at 1% confidence level and to 63 at 2.5% level. We proceeded as follows: For a given trading day, we used the 252 previous trading days (one year) to compute the VaR. The EWMA seed is the first squared daily return of that previous year. We considered a daily update of the VaR, consistently with the methodology previously used to produce figures 1 to 3. FRTB provides a monthly cap regarding the update of datasets (and the recalibration) of internal risk models. A frequent update is likely to reduce the number of VaR exceptions, but will be more difficult to handle due to increased IT, operations costs and the management of risk limits.

We then remark that for quite different values of decay factor, ranging from 0.8 to 0.97, we get quite reasonable number of VaR exceptions. However, if we focus on a stressed sub-period (January 2007 - January 2011), the number of 1% VaR exceptions is equal to 8 for a decay factor equal to 0.8 and increases up to 13 for a decay factor equal to 0.97. Smaller decay factors imply prompter VaR increases when volatility rises and better behaviour during stressed periods.

Since the maximum number of permitted 1% VaR exceptions over past year, for a single desk, is 12, regulatory prescriptions do not seem binding. However, given that stressed periods might be associated with large market moves, improper accounting of negative convexity, i.e. using Taylor expansions in lieu of full revaluation of options exposures, plus the increased occurrence of operational errors might have an extra material impact of the number of VaR exceptions. The backtesting requirement should not thus be dismissed and (internally) assessed over a stressed period, on top of day to day monitoring, in order to assess the resilience of a risk model.

VaR Model/ # of VaR exceptions	$VaR_{1\%}$	$VaR_{2.5\%}$	$(2007-2010)$ $VaR_{1\%}$	$(2007-2010)$ $VaR_{2.5\%}$
VWHS using $\lambda = 0.80$	28	68	8	23
VWHS using $\lambda = 0.85$	26	67	8	21
VWHS using $\lambda = 0.90$	22	66	8	24
VWHS using $\lambda = 0.94$	26	68	10	27
VWHS using $\lambda = 0.97$	24	70	13	28
VWHS using VIX	74	103	35	43
Plain HS (one year window: 2004)	40	89	23	46
Plain HS (two years window: 2003-2004)	45	80	32	49

Table 4: VaR exceptions using EWMA model with different decay factors, VIX approach, and simple HS model for a period of 10 years of VaR exceptions calculation (2005-2014).

We also remark that plain historical simulation may lead to poor backtesting results. According to the EBA study already mentioned, 58% of surveyed banks, with agreed internal models, were using a one-year estimation period, already conforming to FRTB prescriptions. 42% used a longer estimation period. We thus used plain historical VaR based on a rolling one-year period as for VWHS approaches. The number of VaR exceptions is above expectations, 40 and 89 at the 99% and the 97.5% levels respectively. Using a (rolling) two years estimation period (last line of Table 4) does not make a big difference. Though this is not a formal statistical procedure, this suggests that VWHS outperforms plain HS, especially at the 1% level, and better conforms with the expected number of VaR exceptions over the very long run, including periods of market stress, such as the great financial crisis or the European sovereign debt crisis. This is consistent with the findings of Kourouma et al (2010) that emphasizes the poor performance of unfrequently updated VaR models during the 2008 crisis. Similar results are provided by Boucher et al. (2014), where plain HS (λ=1) provides poor results under stress (see also O'Brien & Szerszen (2017)).

While the VIX is acknowledged for its forward-looking properties (see Christoffersen and Mazzotta (2005)), it performs quite poorly when used in the VWHS scheme and is associated with a very large number of VaR exceptions. We might have improved filtering with VIX by appropriate scaling, but this is beyond the scope of our research.

It is of interest that the Basel Committee considered two confidence intervals, 97.5% and 99%, aiming at controlling the accuracy of the tail distribution. Hurlin & Tokpavi (2006), Pérignon & Smith (2008), Leccadito, Boffelli, & Urga (2014), Colletaz et al. (2016) have also advocated the use of different confidence internals, especially due to the capital metrics shift from Value at Risk to Expected Shortfall. Emmer et al. (2015) suggests considering VaR at different confidence levels for backtesting Expected Shortfall, Kratz et al. (2016) further discuss the idea in the FRTB context. It is worth noting that, while the base time horizon for Basel compliant Expected Shortfall is 10 days, the backtesting toolbox (VaR at 97.5% and 99%, p-values) involve a single trading day time horizon.

In the same vein, the Basel Committee quantitative impact studies and the 2013 US regulatory rules prescribe reporting of p-values at desk level. We remind that if F_t is the (predicted) conditional distribution of returns, the p-values are $F_t(r_{t+1})$. Though no formal tests are required, this clearly points out to the PIT (Probability Integral Transform) adequacy tests aimed at checking whether the loss distribution (instead of a single quantile) is well predicted: if F_t is well-specified then, $F_t(r_{t+1}) \sim U$ where U is uniformly distributed over $(0,1)$ leading to formal statistical tests (Crnkovic and Drachman (1995), Diebold et al. (1997), Berkowitz (2001)). We also refer to McNeil (2016) for recent developments in the FRTB backtesting context.

A qualitative inspection of Figure 4, where a decay factor of 0.8 has been used, shows that the QQ plot is well aligned with the bisector line, which is rather reinsuring. However, Figure 5 tempers this outcome. Using a higher decay factor of 0.94 also provides reasonable outcome, consistently with the lack of power of statistical back-tests. For decay factor equal to 0.94 (respectively 0.8), the number of VaR exceptions at 1% level is equal to 26 (respectively 28), while the expected value is 26. For decay factor equal to 0.94 (respectively 0.8), the number of VaR exceptions between the 1% and 2.5% levels is equal to 42 (respectively 40), while the expected value is 41. The figures correspond to the 10 year period starting in 2005.

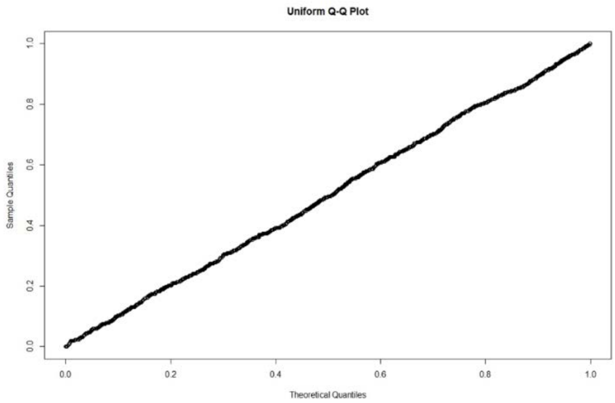

Figure 4: QQ plot for p-values computed under VWHS with decay factor =0.8.

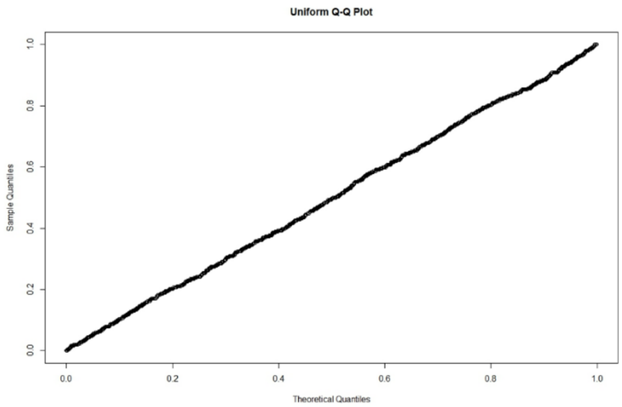

Figure 5: QQ plot for p-values computed under VWHS with decay factor =0.94.

The BIS monitoring exercises also require reporting banks to provide one-day horizon 97.5% values of expected shortfall. Figure 6 provides the ratio the one-day horizon ES to the one-day 99% VaR.

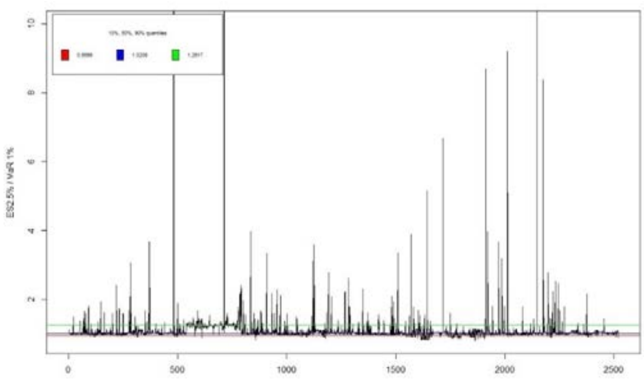

Figure 6: Ratio of 97.5% one-day ES to 99% one-day VaR.

Most of trading days the ratio is close to one, but with quite a number of days where ES is far much larger than VaR. This echoes the debate about the robustness of the ES and the significance of such patterns (Cont at al (2010), Emmer et al (2015)). Though restricted to one-day horizons, Figure 6 suggests that a stressed expected shortfall is likely to be more inflated (as compared with median expected shortfall) than a stressed VaR.

4.2 Dynamic Behaviour of VWHS under Different Decay Factors

As suggested above, resilience of a risk model should be assessed during stress periods, with a concern about clustered VaR exceptions. As outlined by O'Brien and Szerszeń (2017), some internal bank models did perform quite poorly during the great financial crisis, while they used to be overly conservative before.

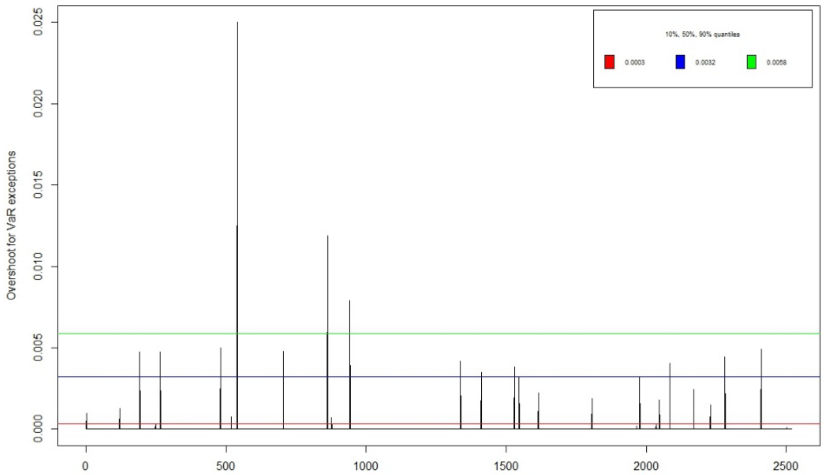

Figure 7: Overshoots for VaR exceptions using VWHS and lambda=0.8 at 1% confidence level.

Clustering of VaR exceptions, i.e. several blows in a row might knock-out bank's capital. Given volatility clustering, risk models that do not react promptly might fail to face extreme events. For this purpose, some tests aim at checking that VaR exceptions are not clustered. This encompasses the duration tests of Christoffersen & Pelletier (2004), Haas (2005), Candelon et al. (2010).

Figure 7 focuses on the location and the magnitude of VaR exceptions over ten years (2005-2014), when using VWHS (decay factor equal to 0.8) on our benchmark S&P500 portfolio. Apart from large overshoots, we can see that VaR exceptions are not concentrated on a specific period, even though rather frequently occurring in groups of two of three exceptions. Even when using a small decay factor and thus a promptly reactive volatility filter, VWHS does not fully comply with the requirement of removing clusters of VaR exceptions.

We consider a formal assessment of VWHS based on the conditional coverage test of Berkowitz et al. (2008) (see also Cenesizoglu and Timmermann (2008), Gaglianone et al. (2012), Dumitrescu et al. (2012), White et al. (2015) for related approaches).

More specifically, Engle and Manganelli (2004) considered the following type of regression as a check of the conditional coverage property, where past VaR exceptions and past values of the VIX index level would be instrumental variables, if a GMM approach had been used.

$$I_t = \alpha_0 + \sum_{i=1}^{I} \alpha_i I_{t-i} + \sum_{j=0}^{K} \beta_j VIX_{t-j} + u_t \tag{4}$$

The VaR model is well specified at $\alpha = 1\%, 2.5\%$ if $\beta_j = 0, \alpha_i = 0, i \geq 1$. Engle and Manganelli (2004) proposed a dynamic quantile test. Berkowitz et al. (2008) assume that the error term follows a logistic distribution, where the log odds are regressed against regressors in Equation (4) above. We followed their approach (implementation is based on the **glm()** function in statistical software R). The outcome depends on the chosen instrumental variables and the associated lags. We chose $I = 4$ as in Engle and Manganelli and the VIX index as an extra instrumental variable as in Berkowitz et al. (2008). We used backward model selection techniques and the Bayesian Information Criteria (BIC) to come up with the appropriate regression model. Results in the case of the 2.5% VaR and a decay factor equal to 0.94 are provided in Table 5. The regressors selected by the backward scheme corresponds to the red cells in Table 5. BIC result in this case is relatively close to the lowest one. The VIX is not selected in either case.

GMM model	$(1\|0)$	$(1\|1)$	$(1\|2)$	$(2\|0)$	$(2\|1)$	$(2\|2)$
BIC	67.18	72.25	69.70	65.07	70.21	67.80
GMM model	$(3\|0)$	$(3\|1)$	$(3\|2)$	$(4\|0)$	$(4\|1)$	$(4\|2)$
BIC	65.07	70.16	67.71	65.07	70.14	67.56
GMM model	$(1,2\|0)$	$(1,2\|1)$	$(1,2\|2)$	$(2,3\|0)$	$(2,3\|1)$	$(2,3\|2)$
BIC	70.33	75.44	73.02	67.86	73.08	70.66
GMM model	$(3,4\|0)$	$(3,4\|1)$	$(3,4\|2)$	$(1,3\|0)$	$(1,3\|1)$	$(1,3\|2)$
BIC	67.86	73.01	70.43	69.97	75.05	72.73
GMM model	$(1,4\|0)$	$(1,4\|1)$	$(1,4\|2)$	$(1,3,4\|0)$	$(1,3,4\|1)$	$(1,3,4\|2)$
BIC	69.97	74.95	72.32	72.42	77.52	75.00

Table 5: BIC for different models. $(1,2|0)$ means that in Equation (9), we considered lags 1 and 2 for I_{t-i} and no term associated with VIX.

As the VaR violations are in practice rare events and the sample size to be considered in practice is not that large to conduct a backtesting procedure, it might raise a problem with correct statistical inference regarding the model accuracy. Therefore, reducing the possibility of having any error types in the statistical inference relying on finite sample inference to accept or reject a statistic test is of importance[20].

Table 6 reports the results based upon a one-year stressed period, from the 2/29/2008 to 2/28/2009: We used a one-year backtesting period as prescribed by the Basel Committee, but chose to focus on a volatile period to assess the resilience of the risk model. As above, we considered a decay factor of 0.94 and computed the 2.5% VWHS VaR. Since backward model selection technique and the use of BIC may lead to different choices of regressors in equation (4), we proceeded either using a single regressor I_{t-3} or two regressors I_{t-3}, I_{t-4} according to the above analysis.

Parameters (one regressor, I_{t-3})	Estimate	Std. Error	z value	$Pr(>\|z\|)$
α_0	-3.8544	0.4519	-8.529	$< 2e-16^{***}$
α_3	2.2450	1.1850	1.894	$0.0582^{\cdot}$
Parameters (two regressors, I_{t-3}, I_{t-4})	Estimate	Std. Error	z value	$Pr(>\|z\|)$
α_0	-4.0561	0.5043	-8.043	$8.77e-16^{***}$
α_3	2.4467	1.2060	2.029	0.0425^{*}
α_4	2.4467	1.2060	2.029	0.0425^{*}

Table 6: Checking the VWHS VaR model. * corresponds to a 5% level.

[20] For instance, Christoffersen and Pelletier (2004), Berkowitz et al. (2008) used a Monte Carlo testing technique proposed by Dufour (2006).

This shows mixed results. When using a single regressor, we do not reject the VWHS 2.5% VaR model, while the converse occurs when using two regressors…

Proceeding in the same way with different values of the decay factor, we come up with statistically significant coefficients in equation (4) when the decay factor is above 0.9, while using a decay factor below 0.9 did not lead to the rejection of VWHS. This is in line with the qualitative insights provided based on the number of VaR exceptions. However, this pleasant feature is no longer valid, if we do not restrict to $I = 4$. Though considering the number of VaR exceptions over stressed periods seems to favour VWHS with rather low values of the decay factor, the outcome of the statistical analysis is blurred.

5. Computation of Capital Metrics under the January 2016 Basel 3 Requirements

While back-testing studies tend to favour lower values of decay factor than the .94 Riskmetrics benchmark and dismiss the use of plain historical simulation, there are well-known drawbacks to using more reactive volatility filters when it comes to risk modelling:

- Risk estimates might become unduly shaky, reacting too promptly to new large market moves. This would make day to day management of risk limits within banks more difficult. Also, this would enhance pro-cyclical features. For this purpose, time-averaging is often implemented, for instance 60 trading day averaging in the Basel 2.5 and Basel 3 frameworks (with slightly different schemes).
- In conjunction with the use of EWMA filter, a low decay factor can result in low estimates of volatility after several calm days, leading to severe underestimation of risk. For this purpose, it can be thought of introducing volatility floors at some stage in the estimation process. Of course, using a volatility floor can lead to a dramatic reduction in the number of VaR exceptions, but conservativeness will then be privileged over accuracy.

Nevertheless, it is important to account for the specificities of capital and backtesting requirements within the Basel 3 framework that can lead to different modelling choices regarding the value of the decay factor than when dealing with initial margin models used bilaterally or by CCPs.

As for capital requirements under the internal models' approach, moving from Basel 2.5 to the FRTB involves switching from a 99% VaR to a 10 days 97.5% Expected Shortfall computed over 12 months. No scaling from a shorter horizon is allowed (see Danielsson and Zigrand (2006) for a discussion), while the use of overlapping observations is explicitly permitted. It is worth noting that the use of historical simulation with overlapping observations coupled with the occurrence of a market crash might lead a repeated use of almost the same 10 days period for the derivation of the ES: the market crash, if severe enough might be included in all the six 10 days periods involved in the 97.5% ES computation. This anchoring effect adds to the well-known issues about robustness of the ES metrics.

As for the derivation of ES under the volatility weighted scheme over a 10 days horizon, we compute the rescaled returns over a 10 days period $(t-h-9, t-h)$ as $\sigma_T \frac{r_{t-h}}{\sigma_{t-h}} + \sigma_{T-1} \frac{r_{t-h-1}}{\sigma_{t-h-1}} + \cdots + \sigma_{T-9} \frac{r_{t-h-9}}{\sigma_{t-h-9}}$, equivalently as $\sigma_T z_{t-h} + \sigma_{T-1} z_{t-h-1} + \cdots + \sigma_{T-9} z_{t-h-9}$[21]. We then sort out these rescaled 10 days returns over a one-year period to derive an empirical expected shortfall. Alexander (2009) provides a forward-looking scheme to the computation of 10 day returns in a FHS context, this

[21] For ease of exposition, we earlier considered relative rather than log-returns. Using daily log-returns makes more sense here since they add-up to provide 10 day-returns, while adding daily relative returns results in an approximation.

requires simulating returns and volatilities by bootstrapping. We also refer to Giannopoulos & Tunaru (2005), Righi & Ceretta (2015), Barone-Adesi et al (2017) for further discussion about implementing the VWHS and FHS for expected shortfall.

As for the computation of volatilities under the EWMA scheme over the one-year period, we proceeded as follows:

- We first computed the average squared return over this one-year period.
- We then used this value as the seed for the EWMA recursion (Equation (2)).

From a capital metrics point of view, all returns required to compute the EWMA seed are known to the modeller.

If the decay factor equals one (plain historical simulation), then the volatility estimates remain equal to the historical volatility, which would not be the case with another seed. Other possible choices might have been made. Some consider an historical volatility computed over a "very long" period as a seed. Another possibility would be taking the first squared return. These seed values are arbitrary and might be unrelated to market conditions over the one year considered period for the computation of the Expected Shortfall. For instance, a very small value of the seed would lead to underestimate the volatilities at the start of the computation year. Therefore, the revol/devol scaling term $\frac{\sigma_T}{\sigma_t}$ is typically associated with quite large values at inception of computational year and gradually decreases to one, its end value. Scaling-up first returns by a multiplicative factor up to 10, means that it will be likely that large negative values will be located at the beginning of the computational year. For a small decay factor, a typical pattern would be using an almost null squared return as a seed, followed by a very large negative return at day 5 or 6. Due to the overlapping 10 days computation windows for the expected shortfall, the risk estimates will then be driven by those two single values.

For values of the decay factor that are close to one, the revol/devol scaling term $\frac{\sigma_T}{\sigma_t}$ will remain quite above one over a large fraction of the computational year, also leading to inflating the expected shortfall. Figure 8 depicts the revol/devol scaling pattern for a given year, associated when using the first squared return as the EWMA seed and λ =0.999. Let us remark that, when $\lambda = 1$ by construction, we have a flat pattern, with revol/devol values equal to 1. But we need to go to higher digits to observe such convergence.

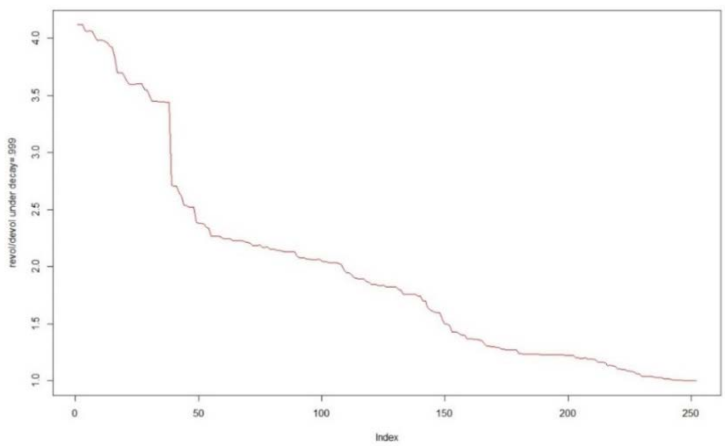

Figure 8: $\frac{\sigma_{252}}{\sigma_t}$ over the period 03 January, 2007 – 02 January, 2008, $\lambda = 0.999$.
Trading day on x – axis. $\sigma_0^2 = r_0^2$.

The endogenous selection magnifies the issue, i.e. the starting day of the one-year stressed period, will be such that the revol/devol scaling pattern exhibits the above feature. An inadvertent, but improper, seed choice leads to VWHS expected shortfall up to seven times higher than plain historical simulation expected shortfall. Occasionally, we end-up with stressed risk measures higher than the nominal exposure: Under plain historical simulation, the expected shortfall is capped by the maximum loss (and thus the nominal exposure), which is not the case under VWHS.

The Basel 3/FRTB capital metrics is fully based on a one-year stress period, as compared to a blend of current and stress periods in the existing Basel 2.5 setting. This one-year stress period under the Basel 3 framework is determined endogenously as the one that will maximise the Expected Shortfall, starting from January 2007 at the latest. It must be updated no less than monthly (we chose a daily update in this paper). Based upon the S&P500 index, we found the following results:

- For decay factor between [0.8, 0.94], the stress window is from Oct 18, 2007 to Oct 16, 2008;
- For decay factor between [0.95, 0.97], the stress window is from Nov 27, 2007 to Nov 24, 2008;
- For decay factor 0.98 the stress window is from Nov 29, 2007 to Nov 26, 2008;
- For decay factor 0.99 the stress window is from Dec 04, 2007 to Dec 02, 2008;
- For decay factor 1 (plain historical simulation), the stress window is from Nov 09, 2007 to Nov 07, 2008.

Though the stress window encompasses the great financial crisis, it does depend upon the chosen decay factor. For a given decay factor, this stress period will remain unchanged until a crisis of larger magnitude occurs. The use of a stress periods kills the procyclical effect usually associated with lower decay factors. Apart from macroprudential views, procyclical capital metrics lead to some practical difficulties for financial institutions. The instability of outcomes jeopardizes the risk management and the setting of risk limits processes.

The Basel Committee new capital metrics involves an averaging of the expected shortfall over the past 60 days. Since the stressed period is unchanged in our case study, this cancels out the averaging effect, which purpose is to deal with time varying exposures. For constant positions, we are left with a single computation: for $\lambda = 0.94$, we need to compute the ES over the period October 18, 2007 to October 16, 2008.

In Figure 9, we provide the revol/devol scaling pattern over that stressed period, associated with $\lambda = 0.94$. Since the seed is set to the average squared return, we do not anymore observe the decreasing pattern of Figure 6. However, the scaling coefficients remain quite larger than one over most of the year. This scheme is associated with an increase in volatility at the end of the period. Thus, large values of $\frac{\sigma_T}{\sigma_t}$ inflate the returns and the expected shortfall as compared with plain historical simulation.

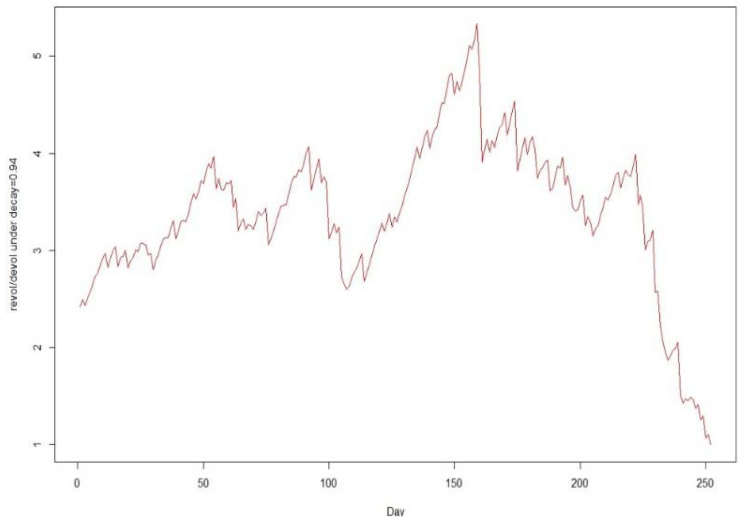

Figure 9: $\frac{\sigma_{252}}{\sigma_t}$ over the period October 18, 2007 to October 16, 2008, $\lambda = 0.94$.
Trading day on x – axis. σ_0^2 is equal to the average squared return over that period.

Thus, unsurprisingly, Figure 10 shows that VWHS leads to higher stressed expected shortfall than when using plain historical simulation.

To ease comparisons, we chose a fixed computation period, 2006-2011 and computed the maximum Expected Shortfall for each decay factor over this historical window. For $\lambda = 0.94$, the VWHS expected shortfall estimate (35%) is roughly 50% higher than using plain historical expected shortfall (22%).

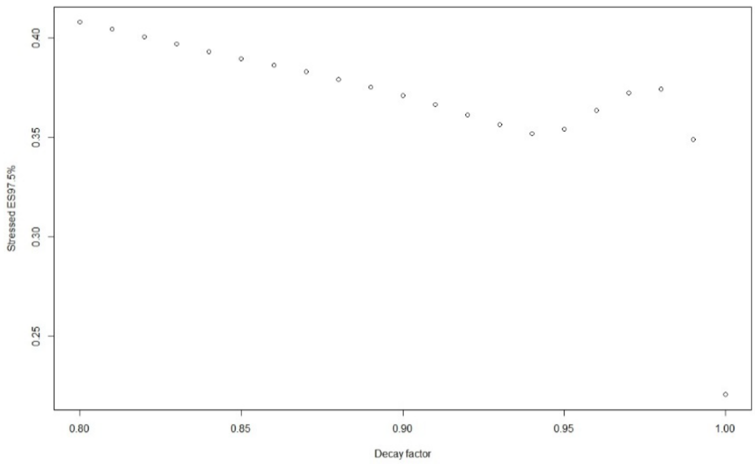

Figure 10: Values of 97.5% expected shortfall computed over a one-year endogenous stressed period for different values of the decay factor (x-axis). Years of calculation: 2006-2011.

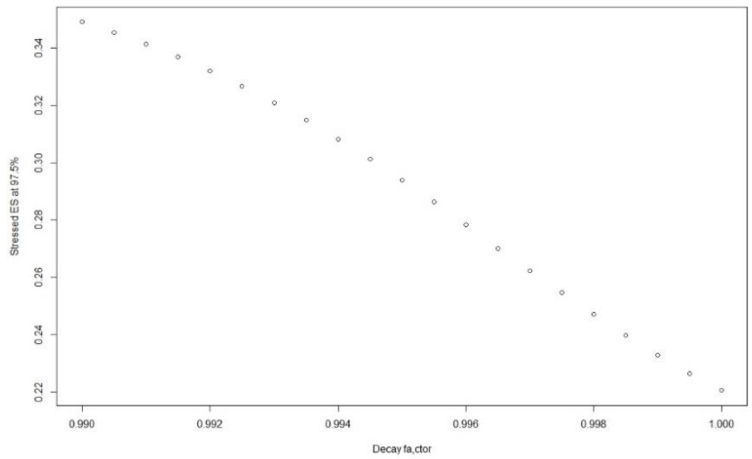

Figure 11: convergence of VWHS to plain historical simulation expected shortfall, decay factor in x-axis. Years of calculation: 2006-2011.

To check the convergence of VWHS expected shortfall ($\lambda < 1$) to plain historical expected shortfall, we provide a zoom for large values of the decay factor in Figure 11. We do observe convergence, but in the provided example, we have a huge dependence of the expected shortfall on the value of the decay factor around 1. We have an expected shortfall of 35% for $\lambda = 0.99$ and of 22% for $\lambda = 1.00$ (plain historical simulation), with an almost linear decrease.

In Figure 12, we provide the revol/devol scaling pattern over the year 2007 for a decay factor equal to 0.9999. The values of $\frac{\sigma_T}{\sigma_t}$ remain quite close to one, consistently with the convergence of VWHS expected shortfall to plain historical expected shortfall.

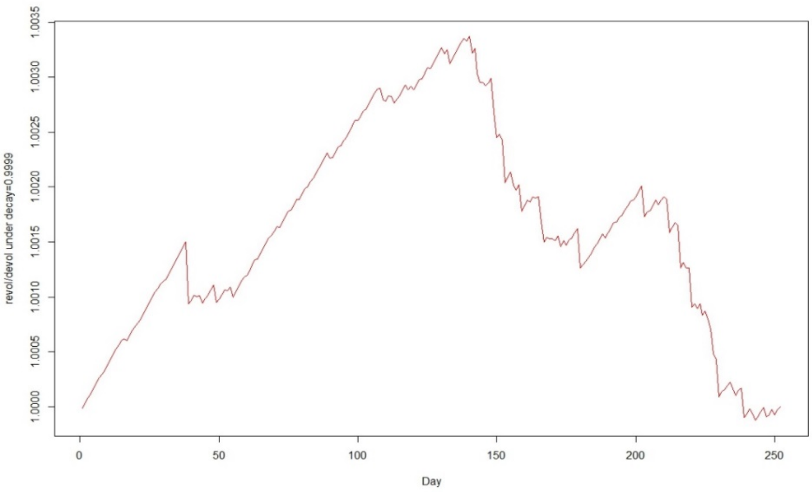

Figure 12: convergence of the scaling pattern to one as decay factor tends to one. $\lambda = 0.9999$. Trading days in x– axis. Year of calculation: 2007.

6. Conclusion

More than twenty years after the disclosure of RiskMetrics and of the 1996 Basel amendment for market risk, quantitative risk models are under scrutiny. They are deemed as overly complex, prone to model error and to tweaking. But where do we stand exactly? How could it be that quantitative risk assessment is considered as altogether prone to model error and to gaming? While the regulation of market risk is experimenting a large overhaul, unveiling the impact of key parameters within the toolbox of risk modellers was the aim of this paper.

Computations based on single day show substantial dispersion of risk estimates, within the class of reasonably well specified models, that do not fail standard back-tests. This variability can be explained by few key parameters that drive the speed at which new market information is incorporated in risk measures. Well-documented difficulties in volatility prediction lead to inherent "statistical model" error. This needs to be understood and accounted for, when assessing the reliability of internal models of market risk measurement.

The 60 days averaging considered under the FRTB framework might be of little help. For the studied benchmark portfolio, the stressed period remains unchanged and the expected shortfall is constant. That is, we are averaging a constant value over 60 days.

Thus, in most cases, the capital charge associated with the Basel 3 FRTB methodology will be acyclic, i.e. will not depend on current market volatility: for a given exposure, unless the current period becomes the stressed period, the capital metrics, i.e. Expected Shortfall computed over a given stressed period will remain constant.

This also suggests an update of risk assessment methodologies currently being used by the Basel Committee or by the European Banking Authority. Apart from the design of hypothetical portfolios, comparing between different bank risk models should involve computations over stressed periods and averaging as now prescribed by the January 2016 Basel rules.

Given the large number of desks potentially involved (up to 100) and the large direct and indirect (reputational, increased supervisory oversight) costs associated with falling-back to standardised computations, managing VaR exceptions is a challenge for large investment banks. Given the specificities the FRTB / Basel III framework for computing and backtesting risk measures, incorporating more quickly market information, i.e. using lower decay factors is a double edge sword. On one hand, better dealing with sudden increase in market volatility lowers the number of VaR exceptions without targeting a series of VaR exceptions in turbulent times and leads to more resilient risk models. On the other hand, this dramatically inflates stressed expected shortfall (even with well suited seeds in the EWMA volatility filter), compared with plain historical simulation. The price of increased reliance is such that it would be unlikely that financial institutions would move from plain historical to volatility weighted/ filtered historical simulation. A volatility floor might be considered as a mean to monitor the revol/devol scheme and mitigate the inflating effects of using stressed expected shortfall. Additionally, it would further lower the number of VaR exceptions and the likelihood of desk ineligibility to internal modelling. Reducing the number of days over which the ES is computed, for instance using monthly computations instead of daily would also mitigate the adverse effects of an endogenous stress period.

There are other challenges when dealing with risk models:

- We restricted to a single risk factor and linear exposure. The number of risk factors involved in practical modelling questions the use of a single decay factor. This is left for future research.
- When it comes to non-linear exposures, banks may need to consider full revaluation. Banks using Monte Carlo approaches to compute risk measures will face increased computer time requirements, since FRTB is more demanding than current regulation. Likewise, plain historical simulation might be favoured over volatility weighted historical simulation.
- The overhaul of existing frameworks concerns both the design of risk methodologies, data (modellable risk factors) and systems (alignment of front office and risk tools, i.e. P&L

attribution tests). While risk factor modellability is out of the scope of this paper, as mentioned above, the use of historical simulation methods might make it easier to fulfil the P&L attribution requirements, even though uncertainties remain regarding the completion of regulatory rules related to the eligibility of internal risk models.

Nevertheless, the issue raised in our paper should be properly addressed. While the use of stressed expected shortfall is more conservative and dramatically reduces procyclicality of capital requirements, it should not provide wrong incentives when it comes to resilience of risk models and jeopardize financial stability.

Bibliography

1) EBA (2017), Results from the 2016 market risk benchmarking exercise, https://www.eba.europa.eu/documents/10180/15947/EBA+Report+results+from+the+2016+market+risk+benchmarking+exercise+-+March+2017.pdf
2) BCBS, *Supervisory Framework For The Use of "Backtesting" in Conjunction With The Internal Models Approach to Market Risk Capital Requirements*, 1996, Available at https://www.bis.org/publ/bcbsc223.pdf
3) Abad, P., & Benito, S. (2013). A detailed comparison of value at risk estimates. *Mathematics and Computers in Simulation*, *94*, 258-276.
4) Alexander, C. (2009). *Market Risk Analysis, Value at Risk Models* (Vol. 4). John Wiley & Sons.
5) Alexander, C., & Sarabia, J. M. (2012). Quantile Uncertainty and Value-at-Risk Model Risk. *Risk Analysis*, *32*(8), 1293-1308.
6) Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2006). Volatility and correlation forecasting. *Handbook of economic forecasting*, *1*, 777-878.
7) Andersen, T. G., Bollerslev, T., Christoffersen, P., & Diebold, F. X. (2007). Practical volatility and correlation modeling for financial market risk management. In *The risks of financial institutions* (pp. 513-548). University of Chicago Press.
8) Angelidis, T., Benos, A., & Degiannakis, S. (2007). A robust VaR model under different time periods and weighting schemes. *Review of Quantitative Finance and Accounting*, *28*(2), 187-201.
9) Barone-Adesi, G., Giannopoulos, K., & Vosper, L. (2017). Estimating the joint tail risk under the filtered historical simulation: An application to the CCP's default and waterfall fund. The European Journal of Finance, 1-16.
10) Berkowitz, J., Christoffersen, P., & Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, *57*(12), 2213-2227.
11) Bignozzi, V., & Tsanakas, A. (2014). Parameter uncertainty and residual estimation risk.
12) Boucher, C. M., Daníelsson, J., Kouontchou, P. S., & Maillet, B. B. (2014). Risk models-at-risk. *Journal of Banking & Finance*, *44*, 72-92.
13) Boucher, C. M., & Maillet, B. B. (2013). Learning by Failing: A Simple VaR Buffer. *Financial Markets, Institutions & Instruments*, *22*(2), 113-127.
14) Boudoukh, J., Richardson, M., & Whitelaw, R. (1998). The best of both worlds. *Risk*, *11*(5), 64-67.
15) Brooks, C., Burke, S. P., Heravi, S., & Persand, G. (2005). Autoregressive conditional kurtosis. *Journal of Financial Econometrics*, *3*(3), 399-421.
16) Candelon, B., Colletaz, G., Hurlin, C., & Tokpavi, S. (2010). Backtesting value-at-risk: a GMM duration-based test. *Journal of Financial Econometrics*, *9*(2), 314-343.
17) Cenesizoglu, T., & Timmermann, A. G. (2008). Is the distribution of stock returns predictable?.
18) Chen, C. W., Gerlach, R., Lin, E. M., & Lee, W. C. W. (2012). Bayesian forecasting for financial risk management, pre and post the global financial crisis. *Journal of Forecasting*, *31*(8), 661-687.
19) Christoffersen, P., & Gonçalves, S. (2005). Estimation risk in financial risk management. *The Journal of Risk*, *7*(3), 1.
20) Christoffersen, P. F., & Diebold, F. X. (2000). How relevant is volatility forecasting for financial risk management?. *The Review of Economics and Statistics*, *82*(1), 12-22.

21) Christoffersen, P., & Mazzotta, S. (2005). The accuracy of density forecasts from foreign exchange options. *Journal of Financial Econometrics*, *3*(4), 578-605.
22) Christoffersen, P., & Pelletier, D. (2004). Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, *2*(1), 84-108.
23) Colletaz, G., Hurlin, C., & Pérignon, C. (2013). The Risk Map: A new tool for validating risk models. *Journal of Banking & Finance*, *37*(10), 3843-3854.
24) Colliard, J. E. (2014). Strategic selection of risk models and bank capital regulation.
25) Cont, R., Deguest, R., & Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative finance*, *10*(6), 593-606.
26) Crotty, J. (2009). Structural causes of the global financial crisis: a critical assessment of the 'new financial architecture'. *Cambridge journal of economics*, *33*(4), 563-580.
27) Danciulescu, C. (2010). Backtesting value-at-risk models: A multivariate approach.
28) Danielsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking & Finance*, *26*(7), 1273-1296.
29) Danielsson, J. (2008). Blame the models. *Journal of Financial Stability*, *4*(4), 321-328.
30) Danielsson, J., Valenzuela, M., & Zer, I. (2016). Learning from history: volatility and financial crises.
31) Danielsson, J., & Zhou, C. (2016). Why risk is so hard to measure.
32) Danielsson, J., & Zigrand, J. P. (2006). On time-scaling of risk and the square-root-of-time rule. *Journal of Banking & Finance*, *30*(10), 2701-2713.
33) Dave, R. D., & Stahl, G. (1998). On the accuracy of VaR estimates based on the variance-covariance approach. *Risk Measurement, Econometrics and Neural Networks*, 189-232.
34) Diebold, F. X., Gunther, T. A., & Tay, A. S. (1997). Evaluating density forecasts.
35) Degiannakis, S., Floros, C., & Livada, A. (2012). Evaluating value-at-risk models before and after the financial crisis of 2008: International evidence. Managerial Finance, 38(4), 436-452.
36) Dowd, K. (2009). Moral hazard and the financial crisis. *Cato J.*, *29*, 141.
37) Dowd, K., Cotter, J., Humphrey, C., & Woods, M. (2008). How unlucky is 25-sigma?. *The Journal of Portfolio Management*, *34*(4), 76-80.
38) Dufour, J. M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics*, 133(2), 443-477.
39) Dumitrescu, E. I., Hurlin, C., & Pham, V. (2012). Backtesting value-at-risk: from dynamic quantile to dynamic binary tests. *Finance*, *33*(1), 79-112.
40) Emmer, S., Kratz, M., & Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, *18*(2), 31-60.
41) Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367-381.
42) Escanciano, J. C., & Olmo, J. (2010). Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics*, *28*(1), 36-51.
43) Escanciano, J. C., & Pei, P. (2012). Pitfalls in backtesting historical simulation VaR models. *Journal of Banking & Finance*, *36*(8), 2233-2244.
44) Fan, J., & Gu, J. (2003). Semiparametric estimation of Value at Risk. The Econometrics Journal, 6(2), 261-290.
45) Francq, C., & Zakoïan, J. M. (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics*, *184*(1), 158-173.
46) Gaglianone, W. P., Lima, L. R., Linton, O., & Smith, D. R. (2012). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics*, *29*(1), 150-160.
47) Giannopoulos, K., & Tunaru, R. (2005). Coherent risk measures under filtered historical simulation. *Journal of Banking & Finance*, *29*(4), 979-996.
48) Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779-1801.
49) González-Rivera, G., Lee, T. H., & Yoldas, E. (2007). Optimality of the RiskMetrics VaR model. *Finance Research Letters*, *4*(3), 137-145.
50) Gourieroux, C., & Zakoïan, J. M. (2013). Estimation-Adjusted VaR. *Econometric Theory*, *29*(4), 735-770.

51) Guermat, C., & Harris, R. D. (2002). Robust conditional variance estimation and value-at-risk. *Journal of Risk*, *4*, 25-42.
52) Gurrola-Perez, P., & Murphy, D. (2015). *Filtered historical simulation Value-at-Risk models and their competitors* (No. 525). Bank of England.
53) Haas, M. (2001). New methods in backtesting. *Financial Engineering Research Center, Bonn*.
54) Haas, M. (2005). Improved duration-based backtesting of value-at-risk. *The Journal of Risk*, *8*(2), 17.
55) Hammoudeh, S., Malik, F., & McAleer, M. (2011). Risk management of precious metals. The Quarterly Review of Economics and Finance, 51(4), 435-441.
56) Harvey, C. R., & Siddique, A. (1999). Autoregressive conditional skewness. *Journal of financial and quantitative analysis*, *34*(04), 465-487.
57) Hull, J., & White, A. (1998). Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk*, *1*(1), 5-19.
58) Hurlin, C., & Tokpavi, S. (2008). Une évaluation des procédures de Backtesting. *Finance*, *29*(1), 53-80.
59) Kourouma, L., Dupre, D., Sanfilippo, G., & Taramasco, O. (2010). Extreme value at risk and expected shortfall during financial crisis.
60) Kratz, M., Lok, Y. H., & McNeil, A. J. (2016). Multinomial VaR Backtests: A simple implicit approach to backtesting expected shortfall.
61) Kuen, T. Y., & Hoong, T. S. (1992). Forecasting volatility in the Singapore stock market. Asia Pacific Journal of Management, 9(1), 1-13.
62) Kuester, K., Mittnik, S., & Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. Journal of Financial Econometrics, 4(1), 53-89.
63) Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. The journal of Derivatives, 3(2), 73-84.
64) Leccadito, A., Boffelli, S., & Urga, G. (2014). Evaluating the accuracy of value-at-risk forecasts: New multilevel tests. *International Journal of Forecasting*, *30*(2), 206-216.
65) Lopez, J. A. (1999). Methods for evaluating value-at-risk estimates. Economic Review-Federal Reserve Bank of San Francisco, (2), 3.
66) Mancini, L., & Trojani, F. (2011). Robust value at risk prediction. Journal of financial econometrics, 9(2), 281-313.
67) Marimoutou, V., Raggad, B., & Trabelsi, A. (2009). Extreme value theory and value at risk: application to oil market. Energy Economics, 31(4), 519-530.
68) McNeil, A. (2016) Backtesting Trading Book Models with Realized p-Values, presentation slides available at https://workspace.imperial.ac.uk/mathematics/Public/Events/London2016.pdf
69) Mehta, A., Neukirchen, M., Pfetsch, S., & Poppensieker, T. (2012). Managing market risk: today and tomorrow. *McKinsey & Company McKinsey Working Papers on Risk*, *32*, 24.
70) Morgan, J. P. (1996). Riskmetrics technical document.
71) Murphy, D., Vasios, M., & Vause, N. (2014). An investigation into the procyclicality of risk-based initial margin models.
72) Nieto, M. R., & Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, *32*(2), 475-501.
73) O'Brien, J., & Szerszeń, P. (2017). An Evaluation of Bank Measures for Market Risk Before, During and After the Financial Crisis. *Journal of Banking & Finance*.
74) Oanea, D. C., & Anghelache, G. (2015). Value at Risk prediction: the failure of RiskMetrics in preventing financial crisis. Evidence from Romanian capital market. Procedia Economics and Finance, 20, 433-442.
75) Pérignon, C., & Smith, D. R. (2010). The level and quality of Value-at-Risk disclosure by commercial banks. *Journal of Banking & Finance*, *34*(2), 362-377.
76) Pritsker, M. (2006). The hidden dangers of historical simulation. *Journal of Banking & Finance*, *30*(2), 561-582.
77) Righi, M. B., & Ceretta, P. S. (2015). A comparison of Expected Shortfall estimation models. Journal of Economics and Business, 78, 14-47.
78) Tse, Y. K. (1991). Stock returns volatility in the Tokyo Stock Exchange. Japan and the World Economy, 3(3), 285-298.

79) White, H., Kim, T. H., & Manganelli, S. (2015). VAR for VaR: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics*, *187*(1), 169-188.
80) Wied, D., Weiß, G. N., & Ziggel, D. (2016). Evaluating Value-at-Risk forecasts: A new set of multivariate backtests. *Journal of Banking & Finance*, *72*, 121-132.
81) Yamai, Y., & Yoshiba, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance*, *29*(4), 997-1015.
82) Ziggel, D., Berens, T., Weiß, G. N., & Wied, D. (2014). A new set of improved Value-at-Risk backtests. *Journal of Banking & Finance*, *48*, 29-41.
83) Zikovic, S., & Aktan, B. (2009). Global financial crisis and VaR performance in emerging markets: A case of EU candidate states-Turkey and Croatia.