FRTB – White Paper

A Statistical Study of the Newly Proposed P&L Attribution Tests

July 2018

One of the key challenges implied by the upcoming regulatory framework for minimum market risk capital requirements, known as FRTB (Fundamental Review of Trading Book), is the P&L attribution (PLA) tests.

PLA tests constitute a game-changer in the way risk engine and models are assessed for effectiveness and accuracy by regulators to base the qualification for the 'internal model approach' for market risk capital. The initial design of these tests, however, has raised concerns from the industry that pointed to a punitive behavior and room for improvement. In this context, the Basel Committee has proposed a completely revamped test design in their March 2018's Consultative Document.

In this paper, we first provide a formal statistical analysis of the performances and the behavior of this newly proposed PLA test as compared to the old one. Our findings suggest that although the revised test design addresses many of the raised concerns, there are still some improvement opportunities, particularly in what relates to the conceptual treatment of hedged portfolios. We also provide a detailed statistical study to orient the selection of the homogeneous PLs distributions test to be retained among the offered alternatives.

Among all the significant changes in market risk capture advocated by the FRTB framework, P&L attribution (hereafter, PLA) tests are the most innovative and critical criteria put forward by the Basel Committee to incentivize the effort to reach well aligned risk and front office systems. Alignment of models and data between the risk engine and the official pricing system (i.e., front-office) is the cornerstone of the FRTB philosophy that emerged in the aftermath of the last market crisis of 2008. To acquire and keep their 'internal model approach' status for a given trading desk, banks must prove on periodic basis that their Risk Theoretical PL (hereafter, RPL) projected from the risk engine is well aligned, in statistical sense, with the official PL stripped from fees and intraday movements, called Hypothetical PL (hereafter, HPL). In other words, what is 'important' for the desk, should be as important for the risk department and vice-versa.

The initial design of these PLA tests (Basel Committee's January 2016 Report of Market Risk Capital Requirement) based on normalized mean and variance ratios of the unexplained PL left between the RPL and HPL, and mostly the variance ratio test, has raised numerous concerns from the industry. The tendency of the test to be punitive in general (excessively high failure rates) and its problematic treatment of hedged portfolios, pointing to a systematic failure of the test by these portfolios, have captured ever since the attention of both industry and regulators. The newly revised PLA test design that was proposed by the Basel Committee in their recently released Consultative Document of March 2018 is an attempt to address those issues and strengthen the conceptual robustness of the PLA test criteria. The new PLA test design leaves the concept of the unexplained PL mean & variance ratios behind and orient the assessment of statistically aligned risk and hypothetical PLs using the combination of two separate tests, each pursuing a specific objective: i) The test of homogenous HPL and RPL distributions, using the two-sample Kolmogorov-Smirnov (hereafter, KS) test or Chi-Square test, and ii) the Spearman correlation of HPL and RPL. On the one hand, the first test of homogenous PLs distributions ensures that the two PLs datasets are generated, in the statistical sense, by similar models. On the other hand, the correlation test enforces the criteria that homogeneity should not be left as the product of randomness in data or a matter of luck, but rather a systematic alignment of HPL and RPL on daily basis in the way they co-variate in response to changes of risk factors or market.

In this White Paper, we examine the newly proposed PLA test design from different angles in order to gain insights about its behavior and ultimately its suitability to assess P&L alignment as intended by the Basel framework. Our focus and analysis are rather of technical nature. We use formal statistical concepts in order to tackle the following key points or questions:

1. To what extent the new PLA test solves the issue of excessive rejection (failure) noticed under the old PLA variance ratio test?
2. Which criteria among the minimum P&L correlation and the homogenous P&L distributions is the most critical in driving the final PLA test outcome? And how these two co-variate?
3. How the new PLA test behaves when dealing with hedged portfolios (perfectly or partially hedged)? And what are the drivers of such outcome?
4. Which test among the KS test and the Chi-Square test is the most robust, according to the formal statistical criteria of test power, to be ideally retained to test for the homogeneity of P&L distributions?
5. How the significance level of the KS and the Chi-Square test could be rationalized using a formal statistical thinking process?

1. Main Findings and Observations

Our main findings worth highlighting are as follows:

1. **Over-rejection:** The newly revised PLA test design addresses the problem of excessive failure rate associated with the old variance ratio test. Evidence provided in this paper clearly shows how the new PLA test behaves in more moderate fashion in rejecting risk models. Our results also indicate that the introduction of the traffic-light rule substantially helps in smoothing out the failure rate in some gray' situations.
2. **Conceptual robustness:** The separation of homogenous PLs distributions testing and minimum PLs correlation condition has a key role in improving the conceptual robustness of the PLA test design. This contradicts with the old PLA variance ratio test, which as we demonstrate later implies a hidden minimum PLs correlation condition that interferes with its second condition over the HPL and RPL variances ratio.
3. **Minimum PLs correlation versus homogenous distributions test:** The minimum PLs correlation level to be met seems more decisive most of the time of the final acceptance/failure outcome under the new PLA test design than the homogenous PLs distributions test. In addition, both PLs correlation and the test of homogenous distributions tend to concur most often in rejecting 'bad' models.
4. **Hedged portfolios:** Like the old PLA test, the newly revised PLA test does not solve the problem of hedged portfolios. Our results, both those produced under the illustrative fixed income example or the ones obtained under the generalized probabilistic model of PLs distributions, show an overwhelming evidence of how the new PLA test systematically rejects hedged portfolios, exactly in the same fashion the old test did. Even as more compelling evidence, hedged portfolios fail the new test when their constituting legs, representing the initiated/outright position and the hedge position, are selected such that they both pass, with a comfortable margin, the PLA test on an individual basis. Overall, this finding should not come at a surprise because PLA tests are conceptually unequipped to deal with hedged portfolios. What we also notice as a problematic finding is that the new PLA test does not only disqualify systematically perfectly hedged portfolios, but also show a clear tendency to punish moderately (or imperfectly) hedged portfolios. We believe that an efficient design to treat hedged portfolios could lie on the way the PLA test is conducted: Instead of conducting the PLA test over the hedged portfolio, one can prove that the constituent legs of the portfolio (i.e., the initiated position and its hedge position) individually pass the PLA test on their own. We recognize however that this treatment might create a room for subjectivity around the way the constituent legs would be selected or defined.
5. **KS or Chi-square test?** Our statistical analysis shows that the KS test is more appropriate than the Chi-Square test to assess the homogeneity of PLs distributions. Besides the fact that the acceptance region of the two tests could be different (with the KS test having a wider region in general), our Monte Carlo study shows that the statistical power of the KS test seems more appropriate. In addition, despite the fact the Basel Committee proposal was prescriptive in terms of the structure of bins to be used to define the range of PLs distributions required by this Chi-Square test, our results show that in general, the test outcome is subjectively impacted by the specification of that structure.
6. **Significance level choice:** The choice of the significance level of the homogenous PLs distributions test can be safely set at the level of 5% commonly used in the statistical literature. Our statistical analysis shows that lowering the significance level from 20% to 5% will not create a substantial loss of the actual test power. Most importantly, we argue that the choice of the significance level should not be made out of an absolute aversion to commit an Error of type II (i.e., not rejecting models when they are 'bad'), since a lower significance level does not imply an actual loss of test power in all scenarios.

2. Conceptual Linkage between the Old and New PLA Tests

To conceptually understand why the newly proposed PLA test design would behave differently from the old PLA test, one needs first to understand the hidden features of the old PLA variance ratio test, which we find critical, but yet not fully understood so far. Most importantly, by knowing the implied conditions behind the old PLA variance ratio test, we can easily establish clear and strong conceptual equivalence between the old and the new PLA tests.

The original PLA variance ratio test (BCBS Report of 2016) consists of satisfying the condition,

(1) $$R_\sigma := \frac{\sigma_U^2}{\sigma_H^2} \leq r_v$$

With σ_U^2 and σ_H^2 are the variances of the UPL (UPL=RPL-HPL) and HPL, respectively, and r_v is a fixed threshold that was originally set at 20%.(a)

Let's denote ρ the correlation between HPL and RPL, and σ_R^2 the RPL variance. Therefore, inequality (1) is equivalent to:

(2) $$R_\sigma := \frac{\sigma_U^2}{\sigma_H^2} = \frac{\sigma_H^2 + \sigma_R^2 - 2\rho\sigma_H\sigma_R}{\sigma_H^2} \leq r_v$$

Note: (a) According to the newly proposed PLA test, HPL and RPL data will be collected over the last 12 months period and tests will be performed on a quarterly basis.

By introducing the useful notation,

$$\lambda := \frac{\sigma_R}{\sigma_H}$$

one can show that the UPL variance ratio condition is simply equivalent to the following inequality constraint on a quadratic function of λ :

(3) $$Q(\lambda) := \lambda^2 - 2\rho\lambda + q \leq 0$$

where $q := (1 - r_v) > 0.$

Hence the following main and general result (notice this is a more generalized result than the one derived by Spinaci *et al*. (2017)).

Result: Since λ>0 and r_v<1 (q>0), the inequality (1) of UPL variance ratio threshold holds if and only if the following two conditions are met:

1. [Real roots or non-empty set of admissible solutions]: Δ= $4(\rho^2 - q) \geq 0,$ meaning that there is a necessary but not sufficient condition on minimum correlation to be met:

$$\rho \geq \dot{\rho}$$

$$\dot{\rho} := \sqrt{1 - r_v},$$

2. $[Q(\lambda) \leq 0]$:

$$\lambda_l \leq \lambda \leq \lambda_u$$

$$\lambda_l = \rho - \sqrt{\rho^2 - q}, \qquad \lambda_u = \rho + \sqrt{\rho^2 - q}$$

where $\lambda_l > 0$ is satisfied (giving $r_v < 1)$ if condition (1) above on minimum correlation is met.

In other words, the UPL variance ratio test (1) can be declined or decomposed into two joint conditions to be met:

— A necessary but not sufficient condition on a minimum correlation of PLs, $\rho \geq \sqrt{1 - r_v}$, to satisfy that ensures a non-empty set of 'success' possibilities,

— Upper and lower boundaries condition on HPL & RPL variances ratio itself: $\lambda_l \leq \lambda \leq \lambda_u$.

Notice that the minimum correlation result $\dot{\rho} = \sqrt{1 - r_v}$ is more general (and does not depend on knowing λ) than the simple minimum correlation threshold $(1 - (r_v/2))$ one can derive from (3) after assuming the equality of variances (i.e., by imposing λ=1).

The formula of the implied minimum RPL & HPL correlation, $\dot{\rho} = \sqrt{1 - r_v}$, to be met by the risk model under the old PLA test, as a function of the PLA variance ratio threshold, r_v, allows us to build the first link between the old PLA test and the newly proposed one. If we ignore the potential deviation between the linear Pearson correlation and the Spearman rank correlation (which is mainly caused by the outliers in the data), one can easily deduce that the minimum correlation threshold of 75% under the new PLA test corresponds to an equivalent minimum UPL variance ratio threshold r_v of 0.438, which is higher than the previous threshold of 0.20.

The departure from the UPL mean and variance ratios to the two-sample KS or Chi-square tests is perhaps the main change introduced under the new PLA test. Obviously, the two-sample KS and Chi-square tests offer a broader and more robust statistical assessment of the homogeneity of HPL and RPL distributions compared to the UPL mean and variance ratios. However, for moderate-tail distributions, the homogeneity of variances goes hand in hand with the distributional homogeneity assessed by these tests.

What constitutes in actual terms the main change underlying the new PLA test design is the departure from the previous situation under the old PLA test where the two conditions of implied minimum correlation $\dot{\rho}$ and the range $[\lambda_l, \lambda_u]$ of homogenous variances (i.e., admissible PLs variances ratio, λ^2) are endogenously dependent on each other's.

Exhibit 1: Conceptual linkage between the old and the new PLA tests

Old PLA Test

- UPL Variance Ratio $R_\sigma := \frac{VaR(UPL)}{VaR(HPL)} \leq r_v$ (r_v = 0.20)
  - Implied Condition of **Minimum (linear) Correlation** of $\dot{\rho} = \sqrt{1 - r_v}$ = 0.894
  - **Implied Boundaries** $[\lambda_l, \lambda_u]$ for the HPL & RPL Variance Ratio $\lambda^2 := \frac{VaR(RPL)}{VaR(HPL)}$

    ➔ Condition of RPL & HPL **Homogeneous Variances**
- UPL Mean Ratio threshold

New PLA Test

- Explicit Condition of **Minimum (rank) Correlation** of 0.75 (lower than the implied minimum correlation of 0.894 under the old PAL test)
- Explicit Test (KS/Chi-Square) of **Homogeneous RPL & HPL Distributions** ➔ Means, Variances, Tails...

Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

3. A Fixed-Income Modelling Example

We illustrate here through a fixed income modelling problem the outcomes of the old PLA test based on the unexplained PL (hereafter, UPL) mean and variance ratios as compared against the newly proposed PLA test based on the Spearman correlation and the KS two-sample test of PLs distributions. We limit the testing for homogeneous PLs distributions to the usage of the KS test, as we will discuss in section 4 the superiority of this test over the alternative Chi-Square test.

We target two main objectives from conducting this analysis:

— Compare the two PLA tests outcomes as applied to the same realistic case in order to assess to which extent the newly proposed PLA test departures from the old test in terms of excessively rejecting aligned PLs systems,

— Compare the outcomes of the two PLA tests in the case of hedged portfolios.

3.1. Risk Modelling Complexity & PLA Tests Outcomes: How often/fast the New PLA Test recognizes a 'Good' Risk Model?

The fixed income modelling problem we propose as an illustrative example consists of fitting the term structure yield curve, to be used for risk pricing, to the data of market yields. Our testing approach consists of controlling ex-ante for the accuracy of the yields curve fit provided by the parametric risk model, so we can study the outcome of PLA tests in function of the complexity of risk modelling task. To do so, we propose the procedure below:

— Step (1): We start with an original data of smoothed market yields defined over standardized tenors over which we apply a slight randomization to introduce the typical noise observed in market quotes.

— Step (2): Without loss of generality, the parametric risk pricing model we use for illustration purposes is the Nelson-Siegel term structure model (Nelson and Siegel (1987)). The curve pillars are reduced to the set of 2Y, 3Y, 5Y, 7Y, 10Y and 20Y points.

— Step (3): The HPL and RPL are computed based on daily variations of both market quotes and fitted risk pricing curves for static positions consisting each of one dollar notional invested in constant maturity coupon bonds with terms matching the curve pillars (i.e., no interpolation was required therefore to avoid introducing additional obstacle to the risk model). The coupon rates are selected to be around the median of yields data (respectively for each tenor) to ensure PLs swings on both sides.

To illustrate our results, four scenarios are considered where the randomness of market yields data (influencing the accuracy of the risk model) has been chosen to monotonically decrease from Scenario (I) to (IV) so that the quality of the parametric risk pricing model increases accordingly.

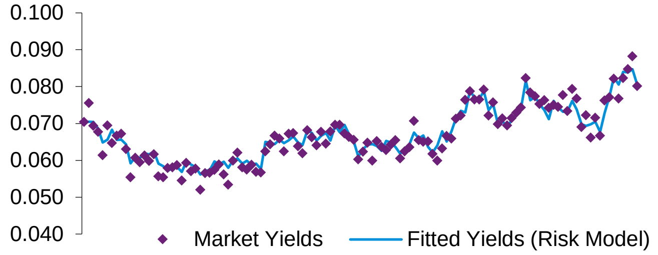
Panel (A) of **Figure 1** illustrates the time series of market yield quotes (underlying HPL) and the fit provided by the parametric curve model for risk pricing (underlying RPL) for the 5Y yield under the four scenarios considered of modelling complexity (as controlled by the intensity of noise in market data points). Panel (B) of **Figure 1** illustrates under the four scenarios the yield curve fit for a given observation date to provide a clear illustration of the overall accuracy of the risk model and how it varies across the considered scenarios.

As we can see, the parametric example examined here provides numerous insights regarding the comparative outcomes of the two PLA tests, which the most important ones are:

— As seen in **Table 1**, while both the old and the new PLA tests respond to the decreased modelling complexity in fitting the risk pricing curve from moving from Scenario I (High difficulty) to the Scenario IV (Low difficulty) by seeing the acceptance rate increasing accordingly, we see that the old PLA test is being punitive in general compared to the newly proposed test. The failure rate under the old PLA test is much higher under Scenarios II and III compared to the new PLA test, where modeling complexity is considered moderate and the risk model accuracy is quite acceptable as shown in **Figure 1**.

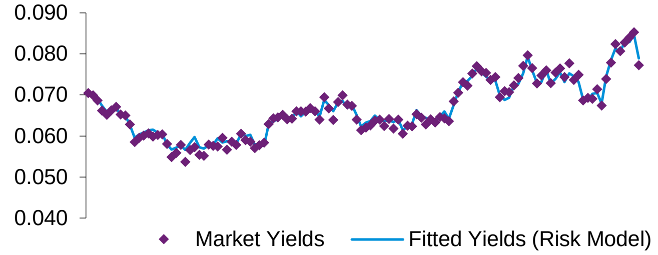
Fig. 1: Illustration of Accuracy of the Risk Model under different Scenarios of Complexity
Panel (A): The Time-Series Snapshot of the 5Y Yield Fit

Scenario (I): High difficulty



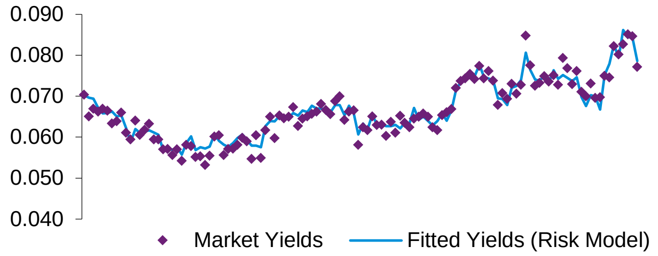
Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

Scenario (II): Medium difficulty



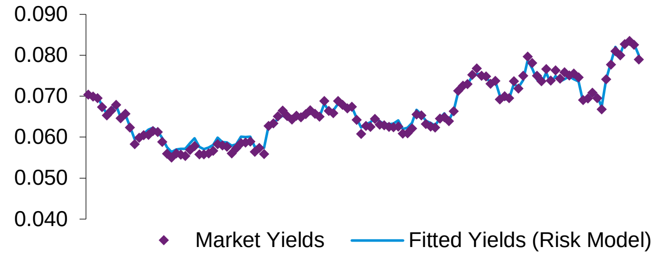
Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

Scenario (III): Medium difficulty



Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

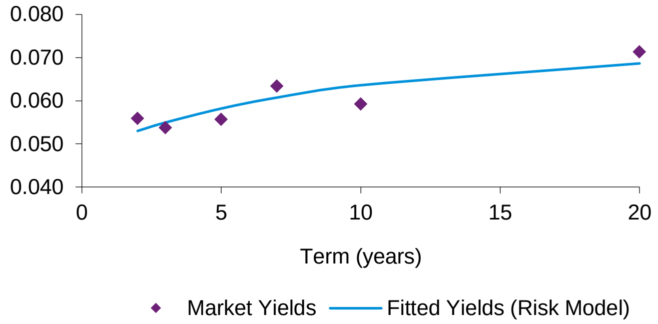
Scenario (IV): Low difficulty



Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

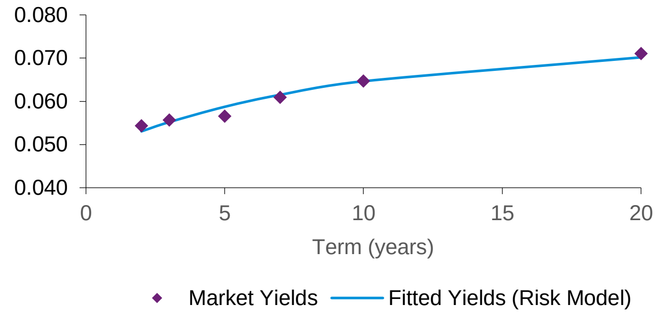
Panel (B): The Yield Curve Fit (for a given observation date)

Scenario (I): High difficulty



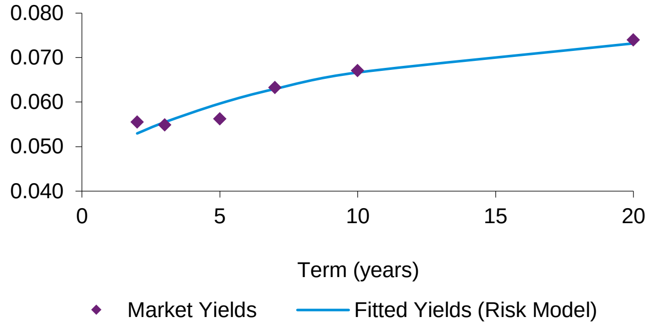
Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

Scenario (II): Medium difficulty



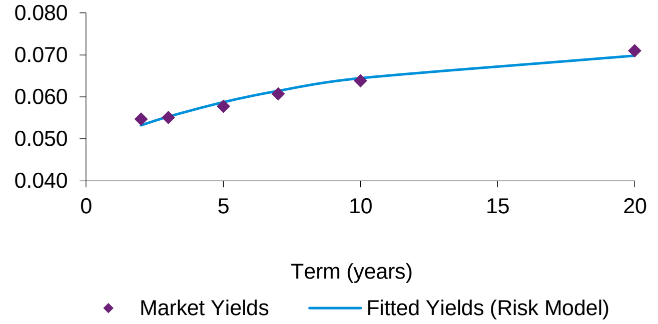
Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

Scenario (III): Medium Difficulty



Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

Scenario (IV): Low Difficulty



Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

— Under the old PLA test, the UPL variance ratio test is the only source of failure of risk models, meanwhile the UPL mean ratio seems accepting all models. This is not surprising knowing the previous analysis made in this direction.

— Under the newly proposed PLA test, the KS test of PLs distributions tends to concur most of the time with the correlation criteria under all the four scenarios considered. It also tends to accept the homogeneity of PLs distributions where the PLs variance ratio (σ_R/σ_H) is closer to 1 rather than the opposite. This is in line with the results derived from the more generalized probabilistic model analyzed in section 4.

Table 1: The Comparative Outcome of the Old and New PLA Tests for the Yield Curve Risk Model Example

		'Old' PLA Test			'New' PLA Test		
Position	std(RPL)/std(HPL)	UPL Mean Ratio (10%) ≤ … ≤ 10%	UPL Variance Ratio … ≤ 20%	Outcome	KS p-value … ≤ 0.20	Spearman Corr. … ≥ 0.75	Outcome
Scenario (I): High Difficulty							
2Y Bond	0,815	0,002	0,284	Fail	0,454	0,840	Pass
3Y Bond	0,757	(0,001)	0,438	Fail	0,454	0,730	Fail
5Y Bond	0,649	(0,001)	0,584	Fail	0,162	0,606	Fail
7Y Bond	0,676	(0,004)	0,509	Fail	0,062	0,671	Fail
10Y Bond	0,663	0,005	0,648	Fail	0,119	0,580	Fail
20Y Bond	0,877	(0,001)	0,154	Pass	0,873	0,910	Pass
Scenario (II): Medium Difficulty							
2Y Bond	0,938	(0,004)	0,180	Pass	0,944	0,869	Pass
3Y Bond	0,823	0,007	0,313	Fail	0,282	0,792	Pass
5Y Bond	0,705	(0,004)	0,467	Fail	0,119	0,710	Fail
7Y Bond	0,751	0,002	0,455	Fail	0,558	0,643	Fail
10Y Bond	0,744	(0,001)	0,343	Fail	0,119	0,775	Fail
20Y Bond	0,964	0,000	0,129	Pass	0,558	0,928	Pass
Scenario (III): Medium Difficulty							
2Y Bond	1,013	0,006	0,174	Pass	0,944	0,872	Pass
3Y Bond	0,855	0,006	0,285	Fail	0,454	0,831	Pass
5Y Bond	0,813	(0,006)	0,268	Fail	0,361	0,813	Pass
7Y Bond	0,765	(0,010)	0,279	Fail	0,558	0,815	Pass
10Y Bond	0,806	(0,001)	0,332	Fail	0,558	0,777	Pass
20Y Bond	0,938	0,007	0,126	Pass	1,000	0,925	Pass
Scenario (IV): Low Difficulty							
2Y Bond	1,034	0,002	0,048	Pass	0,998	0,958	Pass
3Y Bond	0,990	0,003	0,083	Pass	0,944	0,933	Pass
5Y Bond	0,878	(0,004)	0,097	Pass	0,873	0,921	Pass
7Y Bond	0,903	(0,001)	0,089	Pass	0,777	0,941	Pass
10Y Bond	0,913	(0,002)	0,071	Pass	0,873	0,956	Pass
20Y Bond	1,085	0,003	0,061	Pass	0,944	0,973	Pass

Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

— The PLs Spearman correlation seems to be slightly more critical than the KS test in accepting/rejecting risk models. Although the cases in which the risk model was rejected do not show very low correlation, the 0.75 minimum correlation threshold seems to be effective in screening models.

— Interestingly, we do not see any case where the new PLA test is rejecting a risk model while the old PLA test did not already reject it. This is an additional confirmation that the new PLA test does not exhibit an excessive tendency for rejection on its own that could have been differed from that of the old test.

3.2. Hedged versus Unhedged Portfolio: The PLA Testing 'Cliff'

Now, we are interested in comparing the outcomes of the old and new PLA tests when dealing with hedged portfolios.
The excessive punitive outcome toward hedged portfolios the old PLA test has been shown, as it was well-established in a previous research by ISDA, is one of the main concerns of the industry and the Basel committee when dealing with the PLA test design.

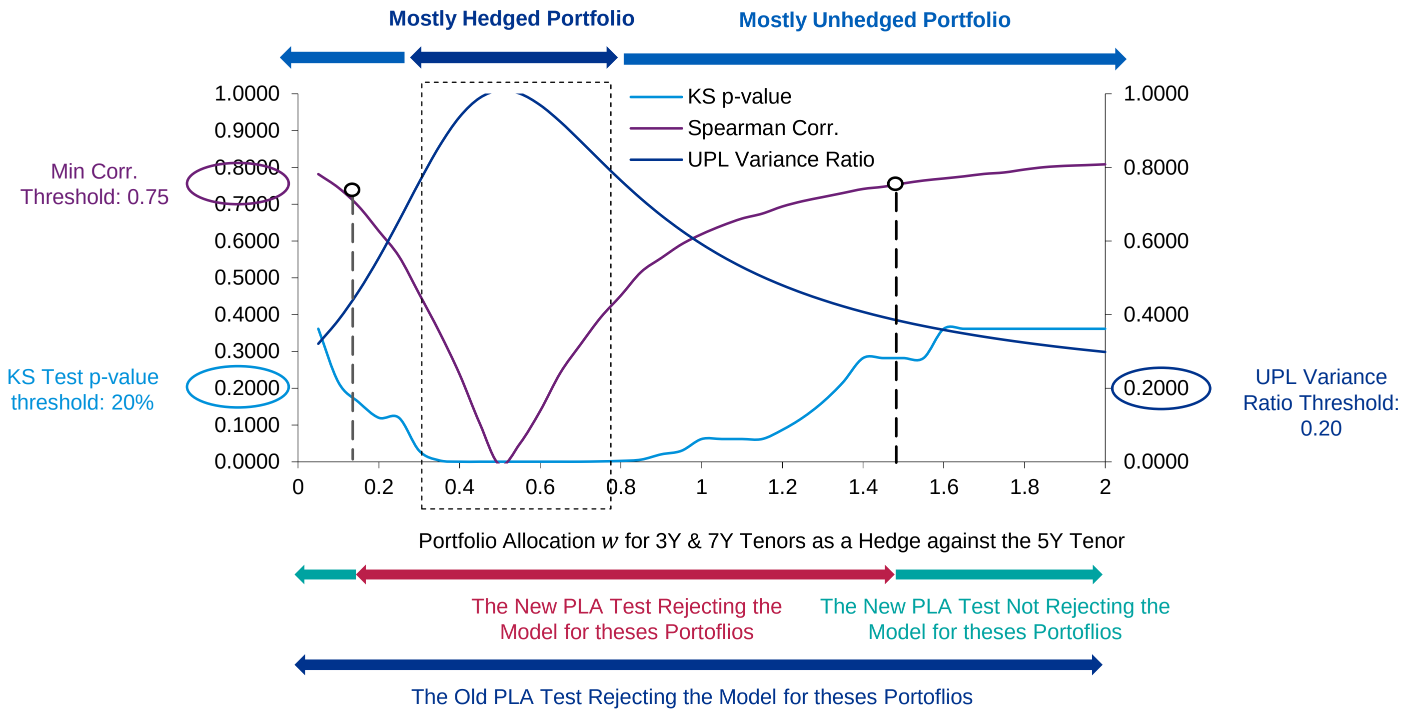
3.2. Hedged versus Unhedged Portfolio: The PLA Testing 'Cliff' (cont.)

We constructed the following simple and realistic case study. First, we start by selecting the data of **Scenario IV**, where **all** risk models as applied to the single positions considered (constant maturity bonds) have passed both old and new PLA tests. The reason is to base our hedge portfolio analysis on the unbiased ground where every single position that could be a potential sub-component (or 'leg') of a broader portfolio (hedged or not) has already individually passed both tests on their own, so that the PLA test outcome at the portfolio level would only capture the pure dynamics of the test at that level. Then, we constructed an intuitive portfolio that mimic commonly used tenor hedging by combining the same single positions analyzed earlier. To illustrate our case, we considered a long position in both 3Y and 7Y tenors to be all together hedging a short position at the 5Y tenor. The portfolio constructed can be therefore parametrized as follows:

$$Portfolio\ (w) :=\ w\ [3Y\ Tenor\ Position + 7Y\ Tenor\ Position] - 5Y\ Tenor\ Position$$

By varying the weight parameter w from zero to values even higher than 1 (as we don't need to restrict the total portfolio DV01 exposure to be bounded), one can construct a set of unhedged and hedged portfolios. Notice that the zone of hedged portfolios corresponds to the area centered around w=1/2, as the two long positions in the 3 and 7 year tenors all together in that case will approximately balance out with the short position in the 5 year tenor.

Fig.2: The Old and New PLA Tests Outcomes for Hedged/Unhedged Portfolio



Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

Figure 2 illustrates the outcome of the old and new PLA tests as applied to the constructed portfolios $\{w_k\}$ by increasing incrementally the parameter w from 0.05 up to 2.

— We observe that both old and new PLA tests concur to significantly reject 'mostly hedged' portfolios (where w is approximately in between 0.30 and 0.80), while the sub-component positions of those portfolios pass, with a large margin, the same tests on their own as shown in **Table 1**. When the new PLA test is considered, we observe that both the KS test and the minimum Spearman correlation condition concur, and in a systematic way, on the rejection of 'mostly hedged' portfolios. Actually, both the KS p-value and the PLs correlation drop in drastic fashion when entering the region of 'mostly hedged' portfolios.

— We also notice that the new PLA test tends to reject most of 'mostly unhedged' portfolios in this parametric example, although again the individual instruments themselves individually pass the tests. With the additional detail to be noticed that the KS test tends to move toward acceptance of those portfolios slightly faster than the minimum correlation test.

— Finally, we also see that the UPL variance ratio test has rejected all the portfolios considered, although this ratio has substantially improved in term of metric value for 'mostly unhedged' portfolios.

It is hard to expect that the results observed here would drastically change under the context of a different example. In fact, additional analysis based on a generalized probabilistic model of PLs distributions also point out to the same observations when it comes to hedged portfolios.

Overall, this points out that an ultimate solution to be considered for hedged portfolios is to apply the PLA tests on the sub-components or 'legs' (i.e., the outright position and its hedge position) of the portfolio. This, however, might come at the cost of allowing for some subjectivity on how these sub-components are stripped out from the total portfolio to form the base over which the PLA tests will be conducted.

4. The KS Test versus the Chi-Square Test

The Basel Committee in their Consultative Document proposed the KS two-sample test and the Chi-square test as two alternatives to be considered to statistically test for the homogeneity of PLs distribution. We provide here a statistical study of the behavior of the two proposed tests. We are particularly interested in comparing the acceptance regions of the two tests when facing the same PLs system and studying their statistical power function as a formal criteria to build an informed judgment about which one of these two tests is the most fit to be retained for the final PLA test design:

4.1. The Acceptance Region of Homogeneous PLs Distributions Tests

We implement here a probabilistic model for PLs distributions to study the acceptance regions of the two homogeneity distributions tests: the KS test and the Chi-Square test. The model we consider assumes that the HPL and RPL samples, $\{HPL_t, RPL_t\}_{t=1,\dots T}$, are generated by the following system:

(4)
$$\begin{cases} HPL_t = Z_{H,t} \\ RPL_t = \lambda Z_{R,t} \end{cases}$$

with $(Z_{H,t}, Z_{R,t})$ are drawn from the joint distribution function $G_{LN}(.)$:

(5)
$$(Z_{H,t}, Z_{R,t}) \rightarrow G_{LN}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)$$

defined as two marginal lognormal distributions (thus allowing for skew and fat-tail effects in PLs), centered around their means and normalized to unit standard-deviations (i.e., Z-scores), and correlated using a Gaussian copula. Which immediately implies:

$$Corr(HPL_t, RPL_t) = \rho, \ \frac{\sigma^2(RPL_t)}{\sigma^2(HPL_t)} = \lambda^2$$

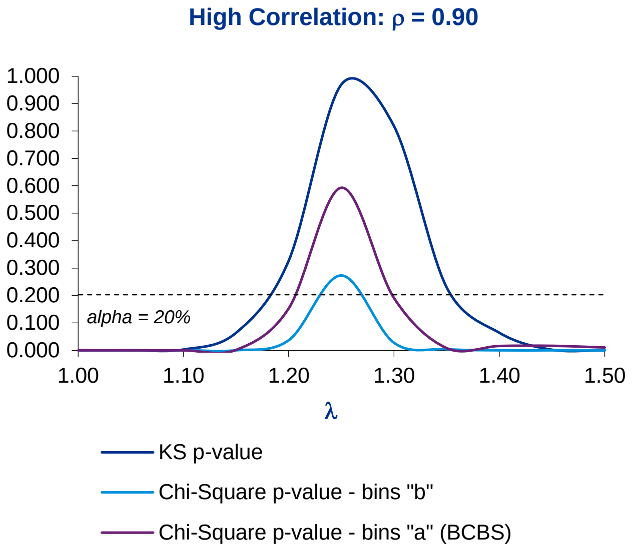
Notice that under this model, one can use the PLs variances ratio, λ^2, as a partial, but yet a strong, indicator of the homogeneity of PLs distributions.

Figure 3 illustrates the p-value of the two tests as function of the PLs variances ratio used as a prior indicator of homogeneity of PLs distributions. Two cases of PLs correlation are considered: a low correlation of 0.60 and a high correlation of 0.90. To execute the Chi-Square test, two structures of 5 bins have been used to highlight the sensitivity of the test to the specification of bins structure (i.e., number of bins and the cut-offs used to position them). Each one differs from the other with respect to the positions of those bins over the range of PL data. The first bins structure '*a*' matches the one prescribed by the BCBS, while the second one, denoted '*b*', uses 5 bins located at different cut-offs by allocating less density or observations on the two extreme segments (tails) of the HPL distribution.

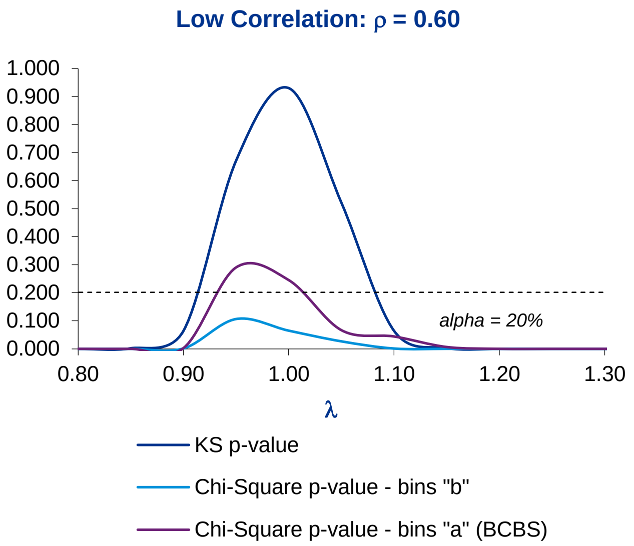
The main observations worth mentioning from these results are:

- The KS test yields a wider acceptance region and the highest p-value compared to the Chi-Square test, although when the p-value exceeds the significance level, this becomes irrelevant. This excess in p-value level is more pronounced under the low correlation scenario.
- The shape of KS p-values is symmetrically positioned over the acceptance region in both low and high correlation scenarios, while the Chi-Square test p-value is slightly skewed under the low correlation scenario, pointing to a potential bias toward lower variances ratios.
- As expected, the level (and the shape) of the Chi-Square test p-value is largely sensitive to the bins structure used to compute the test statistic. Its acceptance region as a result is impacted by the choice of the bins structure. We see this even more pronounced under the low correlation scenario. Moreover, the bins structure suggested in the BCBS document yields a wider region of acceptance in both low and high correlation scenarios than the alternative bins structure used as benchmark.

Fig. 3: The p-values of the KS and Chi-Square Tests of Homogenous PLs Distributions



Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.



Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

4.2. The Compared Statistical Power of Homogeneous PLs Distributions Tests

We study here the statistical power of the homogenous distributions tests to guide the choice between the KS and Chi-Square tests. In general, any null hypothesis testing problem admits two types of errors. The 'Error of Type I' is the error of rejecting the null H_0 where it is true, while the 'Error of Type II' (which is the 'model risk' regulators are the most concerned about) is the error of not rejecting the null H_0 where the alternative H_1 is true. The power function π(α) of the test at the significance level α is defined as:

$$\pi(\alpha) := \Pr_\alpha[\text{rejecting } H_0 | H_1 \text{ is true}] = 1 - \text{Error of Type II}(\alpha)$$

The analytical expression of the power function of the KS and Chi-square tests of homogenous distributions is undefined (or unknown). However, it is possible to infer the power of these tests using Monte Carlo simulations technique. To do so, we consider a simplified example of PLs system under which we can implement an equivalent statistical test that achieves the same null testing, but for which the power function is analytically known.

The simplified example of daily PLs system $\{HPL_t, RPL_t\}_{t=1,\dots T}$, is described as follows:

$$\textbf{(6)} \quad \begin{cases} HPL_t & = F_{H,t} \\ RPL_t & = F_{R,t} + \mu \end{cases}$$

where μ is a constant representing a 'mean shift' or bias parameter and $F_{H,t}$ and $F_{R,t}$ are perfectly uncorrelated variables that follow the joint standard normal distribution,

$$\textbf{(7)} \quad (F_{H,t}, F_{R,t}) \to \mathrm{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$$

As we can see, HPL and RPL in this example are perfectly uncorrelated, since PLs correlation is not our focus in this analysis. We are only interested by the homogeneity of two distributions. The mean parameter μ representing the bias of the RPL model, as compared to the HPL, is the only source here responsible for discarding or approving the homogeneity of PLs distributions.

The Monte Carlo study consists of generating N times the HPL and RPL samples and computing, over the N resulting outcomes of conducting the KS and Chi-Square tests, the empirical frequency $f(\alpha)$ of rejecting the null at a given significance level α for each of the KS and Chi-Square tests. To stress the power of the tests, this empirical frequency $f(\alpha)$is analyzed as a function of the bias parameter, μ.

Notice that under this simplified example of PLs system, the null of homogenous distributions reduces to testing the equality of means of two normal distributions with equal variances, for which the known Student t-test is well suited. Namely, by defining the two-sided t-test of equal means,

$$\begin{cases} H_0\text{: mean}(RPL_t) = \text{mean}(HPL_t) \\ H_1\text{: mean}(RPL_t) \neq \text{mean}(HPL_t) \end{cases}$$

the power function of this t-test is known (see DeGroot and Schervish (2012)), and has the following analytical form:

$$\textbf{(8.i)} \quad \pi_{Student-t}(\alpha|\mu) = \psi_d(-c(\alpha)|\delta(\mu)) + 1 - \psi_d(c(\alpha)|\delta(\mu))$$

$$\textbf{(8.ii)} \quad c(\alpha) = \phi_d^{-1}(1 - \alpha/2)$$

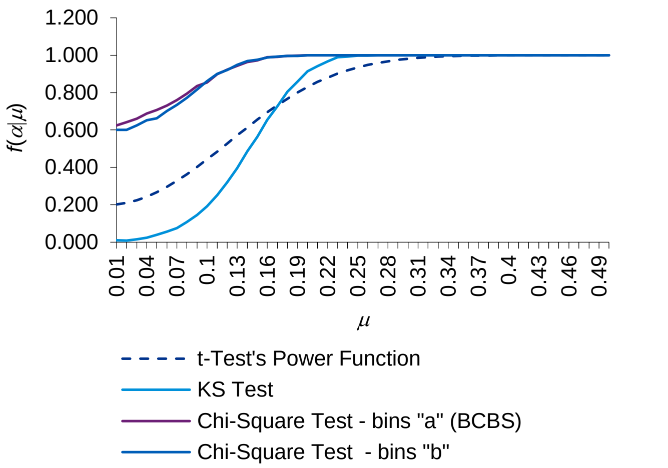
with:

— $\psi_d(.|\delta)$ denotes the c.d.f. of the Non-central Student-t distribution with d degrees of freedom and a noncentrality parameter δ.

— $\phi_d^{-1}(.)$denotes the inverse of the regular (central) Student-t distribution with d degrees of freedom,

and where in our example, we have $d = 2(T-1)$ and $\delta(\mu) = \mu/\sqrt{2/T}$, with T is the size of 12 month-daily PLs samples.

Fig. 4: The Empirical Frequency $f(\alpha|\mu)$of Rejecting the Null as a function of the Bias Parameter μ Compared to the *t*-Test's Power Function $\pi_{Student-t}(\alpha)$



Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

Figure 4 shows the empirical frequency $f(\alpha|\mu)$ we obtain, using $N = 10{,}000$ samples, by varying the bias parameter μ from 0 to 0.50 using the significance level $\alpha = 0.20$ proposed by the Basel Committee. The same two specifications of the Chi-Square test studied earlier with two different structures of 5 bins are implemented here. We also compare in the same plot the empirical functions $f(\alpha|\mu)$ of the KS and Chi-Square tests against the power function $\pi_{Student-t}(\alpha|\mu)$ of the t-test used as a benchmark. We observe that the KS test offers the closest empirical power function $f(\alpha|\mu)$ to the Student t-test power function. By referring to the t-test as a robust benchmark in this case, we can judge that the KS empirical function $f(\alpha|\mu)$ does not indicate a permissive acceptance of the null as it might seem, but rather the manifestation of a balanced test power close to the one exhibited by the t-test. In contrast, and by using the same comparison against the t-test benchmark, the Chi-Square test seems to exhibit an over-rejection of the null.

5. The Choice of Significant Level

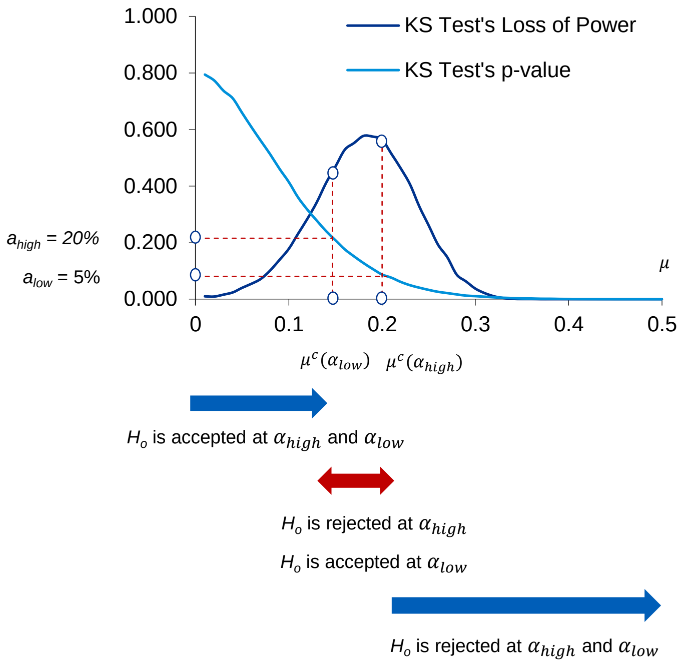
Finally, to examine the choice of the significance level α for the homogeneity test, we will place ourselves under the same simplified PLs system example (6)-(7) examined earlier. Also, given the previous results establishing its superiority, we only examine here the KS test.

Now, to gain additional insights, we need to introduce the metric of loss of test power we define as the loss of the power function due to the shift from a higher significance level, α_{high}, to a lower level,, α_{low}:

(9) $$\mathcal{L}(\alpha_{high}, \alpha_{low}|\mu) \coloneqq \pi(\alpha_{high}|\mu) - \pi(\alpha_{low}|\mu) \geq 0$$

Figure 5 illustrates the loss of test power of the KS test when shifting from 20% to 5% of significance level. In the case of the KS test, the loss function is computed based on the differential of the empirical functions, $f(\alpha|\mu)$. The average p-value of the KS test, as function of μ, over the N generated samples is also displayed.

Fig. 5: The Loss function of Test Power and p-value as a function of the Bias Parameter μ



Source: FRTB — White Paper. A Statistical Study of the Newly Proposed P&L Attribution Tests. 2018.

We see that the loss function of the KS test reaches its maximum level when the bias μ is of a medium level, not too low nor too high. At this region, the Error of Type II, the regulators are the most concerned about, is the highest. On the other hand, as expected, the average p-value of the test over the repeated random sampling experiments, is monotonically decreasing in function of the bias μ. Knowing that the null is rejected anyway when the p-value is below α, the actual loss of test power is bounded between the levels reached at the critical points $\mu^c(\alpha_{low})$ and $\mu^c(\alpha_{high})$ verifying,

$$p - value(\mu^c(\alpha_k))\ ) = \alpha_k, k = "low", "high"$$

while it drops to zero outside these boundaries.

By weighting the p-value of the test into the trade-off, the actual loss of test power is bounded by the critical bias region $[\mu^c(\alpha_{high}), \mu^c(\alpha_{low})]$, where the test outcome diverges depending in which significance level was chosen, while it goes to zero outside it (because in that case the two significance levels lead to the same test outcome). The level of that actual loss of test power over the critical bias region is not significantly higher from the level of the theoretical loss of test power $\mathcal{L}(.)$ outside that region. In addition, the faster the p-value drops when μ increases, the narrower becomes the divergence region, $[\mu^c(\alpha_{high}), \mu^c(\alpha_{low})]$

This analysis shows that the shift from the proposed 20% significance level to the level of 5%, which is most commonly used in the statistical literature and in similar statistical testing exercises employed to validate Basel capital models under different risk classes, does not necessarily introduce a significant loss of test power in actual terms.

Although the degree of aversion to Error of Type II (which is still a risk over the divergence region) is the ultimate driver in this type of decision-making problem, we show how the choice of the significance level should not be made out of an absolute aversion to commit an Error of type II (i.e., not rejecting models when they are ‘bad’), since a lower significance level does not imply an actual loss of test power in all scenarios.

6. Concluding Remarks

We provided in this paper a formal analysis comparing the behavior and the drivers of the newly proposed PLA test design, recently released by the BCBS in their Consultative Document of March 2018, to the old PLA test originally proposed in the BCBS Report of January 2016. The evidence gathered throughout the diverse analysis we conducted points out to major improvements achieved by the new PLA test in terms of conceptual soundness and resolving the excessive failure rates attributed to the old PLA test. We also conducted a formal statistical study using the concept of test power to guide the choice between the two homogenous PLs distributions tests proposed by the BCBS, the KS test and the Chi-Square test, and rationalize the significance level for this testing. Our analysis, however, shows a clear evidence that like the old PLA test, the newly proposed PLA test does not solve the problem of systematic failure of PLA by hedged portfolios. We argue that an ultimate solution would consist in the way the PLA testing exercise is conducted for hedged portfolios rather than the test design itself.

References

1. Basel Committee on Banking Supervision, Consultative Document, “Revisions to the Minimum capital requirements for market risk”, Bank of International Settlements, March 2018.
2. Basel Committee on Banking Supervision, Report, “Minimum capital requirement for market risk”, Bank of International Settlements, January 2016.
3. DeGroot, M., and M. J. Schervish, Probability and Statistics, Fourth Edition, Addison-Wesley, 2012.
4. Nelson, C. R., A. F. Siegel, “Parsimonious Modeling of Yield Curve”, Journal of Business, 1987, vol. 60, no. 4, pp.473-489.
5. Spinaci, M., M. Benigno, A. Fraquelli, and A. Montoro, “The FRTB’s P&L attribution-based eligibility test: an alternative proposal”, Risk Magazine, Dec 2017.

Contact us

Mohamed Mokhtari
Partner
KPMG in Canada
E: mmokhtari@kpmg.ca

Robert Smith
Partner
KPMG in the UK
E: robert.smith@kpmg.co.uk

Ridha Mahfoudhi
Director
KPMG in Canada
E: rmahfoudhi@kpmg.ca

Laurent Duvivier
Manager
KPMG in the UK
E: laurent.duvivier@kpmg.co.uk