Here it “GOES” again: Reviewing the NAIC’s second industry field test of the Generator of Economic Scenarios

Collin Davidson
Parker Henley
Nick Lee
Zohair Motiwalla

Executive Summary

The National Association of Insurance Commissioners (NAIC) continues moving forward with its efforts to reform the Generator of Economic Scenarios (GOES) used in principle-based reserving (PBR) in the United States. Most recently, this entailed a second industry field test, supported by published model office results to understand potential impacts to statutory requirements on business subject to PBR, notably VM-21 reserves and C-3 Phase II capital for variable annuities (VAs). The analysis presented in this paper leverages a model and framework published in 2023 following the first industry field test to build upon current understanding of the proposed scenarios. These calculations are intended to be illustrative of a prototypical VA block and may not be representative of a given company’s situation but aim to provide valuable commentary around risk management implications of the proposed scenarios. Utilizing the economic scenarios presented in the NAIC’s second field test, this analysis looks at VM-21 reserves and C-3 Phase II capital impacts across a range of capital market sensitivities and provides a detailed view of tail risk measures. The paper also highlights key statistical observations and discusses various approaches to the modeling of future hedge strategies that could impact the materiality of the proposed scenarios. Collectively, these analyses document themes and decision points companies should monitor as the NAIC moves closer to the eventual implementation of GOES reform.

Reforming GOES: Potential impacts to statutory reserves

GOES reform has been an NAIC priority for several years, and the second industry field test represents a critical step toward implementation of a new scenario generator in PBR valuation. The first paper in our series, “NAIC Economic Scenario Reform: A Model for VM-21 Impact Analysis,” published in July 2023, covers the background of the GOES reforms and an in-depth analysis of the scenarios provided in the first field test. We encourage the reader to review the introduction of that paper for further background.

This paper continues the focus on potential impacts to statutory reserves and capital for VAs, intended for an audience interested in a mix of technical detail and business implications of GOES reform. There still remains some uncertainty around the specifications for the final adopted version of GOES and when it will be formally implemented.[1] However, signaling from the NAIC and the potential materiality of the new scenarios underscores their importance, making a strong case for the type of analysis presented in this paper. In June 2024, NAIC-supported model office results were published by Oliver Wyman, offering a similar illustrative VA impact analysis. We believe this paper complements the model office testing, but also offers additional insights, all the while maintaining a consistent model and approach with the first paper in this series.

1. As of the date of publication, public NAIC comments have suggested a January 1, 2026, adoption at the earliest, consistent with the expected earliest (optional) adoption date for VM-22, the PBR framework for non-variable annuities.

In this paper, we set the stage with comparative statistics between the Academy's Interest Rate Generator (AIRG) and Conning's GEMS scenarios that form the basis of the second GOES industry field test. Unless otherwise stated, the comparisons throughout the paper will focus on the scenarios provided using the starting yield curve of December 31, 2023, which corresponds to the GEMS "Field Test 1" set. Then we dive into the impact analysis, reviewing key metric outputs using the prototypical VA PBR model that we introduced in our earlier paper, and investigating two sensitivities relating to the choice of a future hedging strategy. We conclude with a summation of themes and some practical considerations to keep in mind as GOES reform continues.

Comparing AIRG and GEMS scenarios

Both the AIRG and GEMS scenarios were calibrated based on a wide range of acceptance criteria. The values below represent only a minor subset of observations comparing the two GOES models and are partially dependent on the level and shape of the December 31, 2023, starting curve used.[2]

In addition to the differences highlighted in the chart in Figure 1, there are several other points of commentary worth noting. Collectively, these observations are based on comparisons of the scenario output and we do not discuss at length the technical complexities of the AIRG or GEMS models that may be underlying these observations.[3]

- **Equity volatility:** Although individual index volatilities are comparable, a measure of blended equity returns produces higher realized volatility in the GEMS model in part due to stronger positive correlation between equity indices.
- **Bond returns:** The GEMS model offers a more sophisticated corporate model that reflects stochastic, mean-reverting credit spreads. As seen, this can lead to higher average returns but also considerably more volatility in bond returns.
- **Rate distribution:** The average, range, and volatility differences in the observed 10-year rate point to a much wider distribution of projected rate scenarios in the GEMS model compared to the AIRG.
- **Curve inversion:** The December 31, 2023, curve has a significant 10-year versus 3-month rate inversion. The statistics highlight how the GEMS model generally maintains the initial curve shape for years longer, but ultimately shifts to a steeper curve shape in later projection years. We also see the frequency of curve inversion in the GEMS model vary more based on the initial curve level, whereas the AIRG has a more stable frequency of inversions.

FIGURE 1: AIRG AND GEMS

BLENDED EQUITY[4]	AIRG	GEMS	GEMS – AIRG
Average Return	10.0%	9.6%	-0.4%
Realized Volatility	17.9%	18.2%	+0.3%
Sharpe Ratio @3% risk-free	39%	36%	-3%

LONG-TERM CORPORATE BOND	AIRG	GEMS	GEMS – AIRG
Average Return	4.6%	5.9%	+1.3%
Realized Volatility	6.6%	12.5%	+5.9%
Sharpe Ratio @3% risk-free	24.5%	22.9%	-1.6%

2. It is possible that analyzing output from both economic scenario generators for a different valuation date could result in different observations.
3. Technical documentation of the GEMS model is available on the NAIC's landing site for the industry field testing, at https://naic.conning.com/scenariofiles.
4. Returns are shown on an annualized basis. Blended equity assumes 40% large cap, 20% international, 20% small cap, and 20% aggressive equity.

FIGURE 1: AIRG AND GEMS (CONTINUED)

INTEREST RATES – 12/31/2023 CURVE	AIRG	GEMS	GEMS – AIRG
Average ultimate 10Y rate	3.33%	4.98%	+1.65%
25th Percentile ultimate 10Y rate	2.52%	2.62%	+0.10%
75th Percentile ultimate 10Y rate	3.88%	6.90%	+3.02%
Average 10Y rate volatility	1.27%	3.26%	+1.99%
Initial 10Y-3M spread	-1.59%	-1.59%	0.00%
Average 10Y-3M spread after 10 years	1.09%	0.67%	-0.42%
Average 10Y-3M spread after 30 years	1.10%	1.21%	0.11%
Frequency of 10Y-3M Inversion	9.3%	21.9%	12.6%
Average Difference when Inverted	0.54%	1.19%	0.65%
Frequency of Negative 3M Rate	0.0%	7.5%	7.5%

Impact analysis

The valuation model used in this analysis is kept consistent across runs used for each scenario set and is intended to reflect assumptions and methodology that are considered illustrative of existing VA business. The table in Figure 2 highlights the chosen base case across a few key dimensions of the illustrative VA business.

FIGURE 2: BASE CASE

DIMENSION	DESCRIPTION
Rider/Guarantee Mix	67% GLWB (with GMDB), 33% GMDB only (step-up feature)
Average Block Age	Mature block, policy duration = 15 GLWB: 25% in-the-money[5]; GMDB: at-the-money 75% of GLWB in withdrawal phase
Fund Mix	70% blended equity, 30% long-term corporate bond
Future Hedging Strategy	Rider-only implicit hedge,[6] with 5% ineffectiveness Final Total Asset Requirement (TAR) assumes 10% E-Factor
Actuarial Assumptions	Approximate VM-21 Standard Projection,[7] with a 5% prudence factor on mortality, lapse, withdrawal efficiency, and expense assumptions.

5. Moneyness is defined on a nominal basis, which is: (Benefit Base / Account Value - 1) for Guaranteed Lifetime Withdrawal Benefit (GLWB). “Benefit Base” refers to the accumulated guaranteed value used to determine the GLWB withdrawal amount and is fixed upon first withdrawal.
6. The implicit hedging approach used assumes a future hedging strategy covering total equity and interest rate risk (or simply “full delta-rho”), based on the rider cash flows (rider charges, GLWB claims, and Guaranteed Minimum Death Benefit [GMDB] claims) as the hedge target. The hedge cost is a risk-neutral valuation based on Milliman Guarantee Index® volatility as of December 31, 2023, assuming the 1-year constant maturity Treasury (CMT) forward curve as the risk-free rate.
7. Based on the assumptions published in the 2025 Valuation Manual. The first paper assumed standard projection assumptions published in prior versions of the valuation manual.

The baseline model runs assumed a $10 billion cash surrender value (CSV) in-force book. This section details the impacts to VM-21 reserves in excess of CSV and after-tax conditional tail expectation (CTE) 98, also in excess of CSV. After-tax CTE98 Total Asset Requirement (TAR) is equivalent to VM-21 reserves plus a target 400% level of C-3 Phase II capital, or simply referred to as "CTE98 TAR."[8] Consistent with the VM-21 instructions, the reported reserve and pre-tax CTE98 TAR results assuming a future hedging strategy and 10% E-Factor are given by the following formula:

"Final" CTE metric = Hedged CTE metric + 10% x max(0, Unhedged CTE metric - Hedged CTE metric)

The initial baseline case reserves across each scenario set, as of December 31, 2023, is shown in Figure 3. Alternatively, Figure 4 presents CTE98 TAR.

FIGURE 3: VM-21 RESERVES

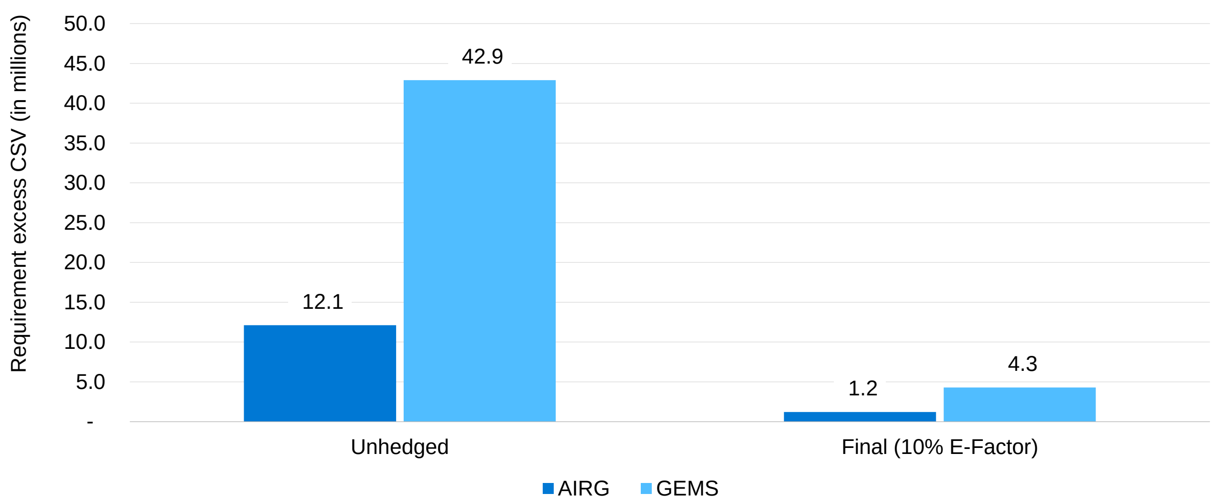
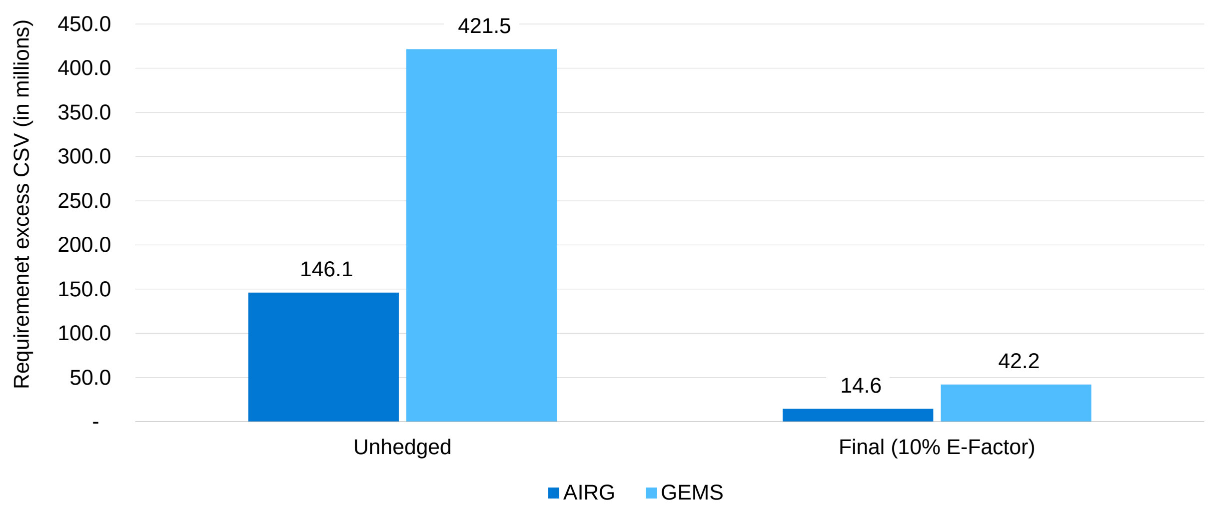


FIGURE 4: CTE98 TAR



8. After-tax CTE98 TAR = VM-21 reserves + (Pretax CTE98 - Pretax CTE70) x (1 - 21%). This implicitly assumes that the additional standard projection amount is zero, and that there is no non-admitted deferred tax asset allocated to the variable annuity line of business in this situation (and so the statutory reserve less tax reserve component is capped at zero).

These results illustrate that, on an unhedged basis, the GEMS scenarios produce materially higher TAR than the AIRG—*totaling +$275 million for the modeled business.* However, with an effective hedge strategy modeled, the difference between the GEMS and AIRG scenarios is dramatically reduced. Still, these results suggest a moderate TAR increase for the VA business: *+$28 million*, which represents a 290% increase in TAR.

It is important to emphasize how the impact of the GEMS scenarios will vary across companies. Within this analysis, the base case business and assumption mix produces a hedged CTE98 that is bound by the CSV floor, so the impact of the GEMS scenarios is directly tied to the amount of unhedged liability cash flow[9] impact contributing to the final TAR. Furthermore, to the extent market conditions change or product features and assumptions vary dramatically from the base case, the materiality of the GEMS scenarios is expected to change.

RESULTS BY CTE LEVEL

To better understand the baseline impact on reserves and capital, we can review the dispersion of results by scenario in the tail of the distribution. Figure 5 plots the scenario-level requirement for each of the worst 300 scenarios (i.e., CTE70), which equals the CSV plus the greatest present value of accumulated deficiency (GPVAD). Note that scenarios plotted below the CSV—represented by the black horizontal line—would be floored at the CSV in the final TAR calculations.

FIGURE 5: SCENARIO-LEVEL REQUIREMENT FOR CTE70 SCENARIOS

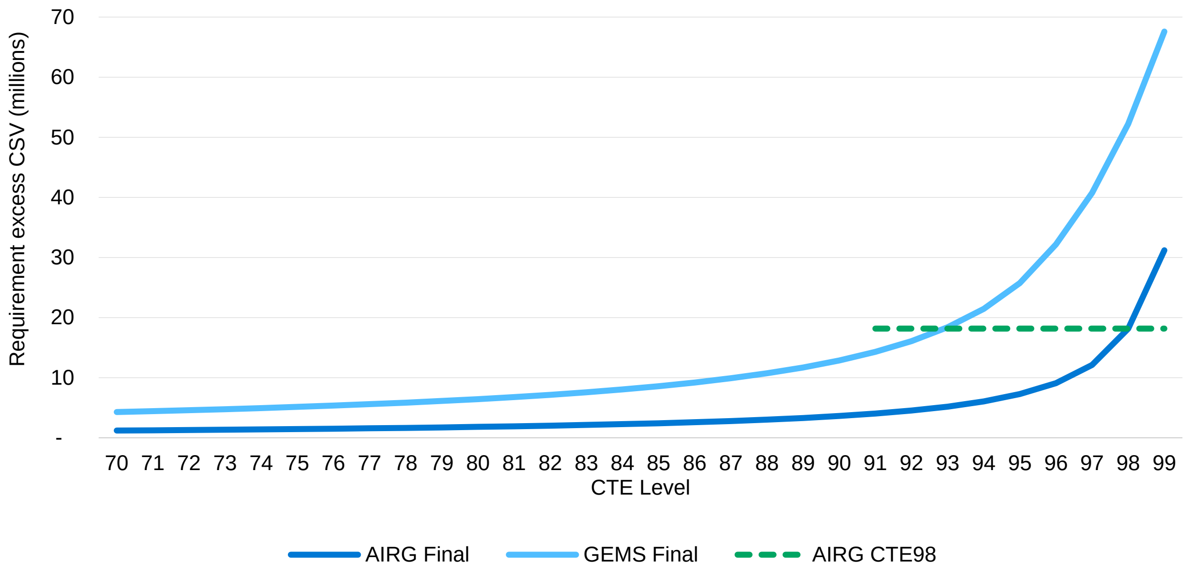
From the chart in Figure 5, we observe:

- GEMS produces a consistently higher requirement by scenario, but this increase is significantly compressed on a hedged basis, consistent with reserve impacts shown in Figure 3.
- GEMS has 39 scenarios (so 3.9% of the total 1,000 set) producing a requirement above the CSV floor, more than twice the 19 scenarios above the floor from the AIRG result.
- Shifting from unhedged to hedged results significantly reduces the dispersion by scenario for both hedged and unhedged, resulting in a remarkably close pattern of results between GEMS and AIRG. Remaining variation across scenarios is driven by the unhedged cash flows and assumed hedge ineffectiveness.

9. This includes cash flows generated by the base contract as well as rider cash flows retained due to assumed hedge ineffectiveness.

As an alternative view, we can look at the calculated CTE level for each scenario set. Figure 6 emphasizes how the GEMS scenario set leads to a much heavier tail than AIRG, where the CTE70 impact is $3 million, but the CTE98 impact is $34 million. In fact, as the horizontal dotted line shows, the AIRG CTE98 level is equivalent to only the *CTE93* level from the GEMS result.

FIGURE 6: FINAL PRETAX CTE RESULTS BY CTE LEVEL



-

MARKET SENSITIVITIES

We also explore the potential impact to market sensitivity profiles between scenario sets. For companies that are interested in hedging their statutory balance sheet to maintain stable risk-based capital (RBC) multiple levels, this can be a challenging task to manage because the PBR mechanics in statutory reporting still exhibit some dynamics that may behave differently from their respective economic hedge target metrics. The situation is also made more complicated by the asymmetry created by the cash surrender value floor. To measure the effects of the GOES on this, we evaluate the level of TAR at risk in “adverse shocks,” as well as some approximate hedging Greeks implied by runs using each scenario set.

For the purpose at hand, we consider the adverse shock as a -25% equity shock, coupled with a significant drop in interest rates. The drop in interest rates is captured by using the low-rate shock scenarios provided (“Field Test 2”), which utilized the March 9, 2020, starting yield curve and amounted to a -334 basis point reduction in the 10-year rate relative to December 31, 2023.

The charts in Figures 7 and 8 show the magnitude of the change in unhedged and hedged TAR from the adverse shock, decomposing the impact between the equity, rate, and cross-effect components.

- In both cases, the GEMS result led to a larger increase in TAR, although the increase in sensitivity was more pronounced on the unhedged basis.
- The equity sensitivity was similar between the AIRG and GEMS results, but the GEMS result was driven by higher rate sensitivity, partially offset by lower unfavorable cross-effects.
- For both AIRG and GEMS results, the sensitivity displayed in the charts is more adverse for the post-hedge TAR. This is primarily driven by the increased interest rate sensitivity due to the implicit hedge shifting the results closer to greater risk-neutral sensitivity. On its own, the -25% equity shock is absorbed by the baseline being partially bound by the CSV floor but is then fully realized when combined with the negative rate shock, contributing to the more adverse cross-effect in post-hedge TAR.

FIGURE 7: ADVERSE SHOCK TAR SENSITIVITY: POST-HEDGE CTE98 TAR

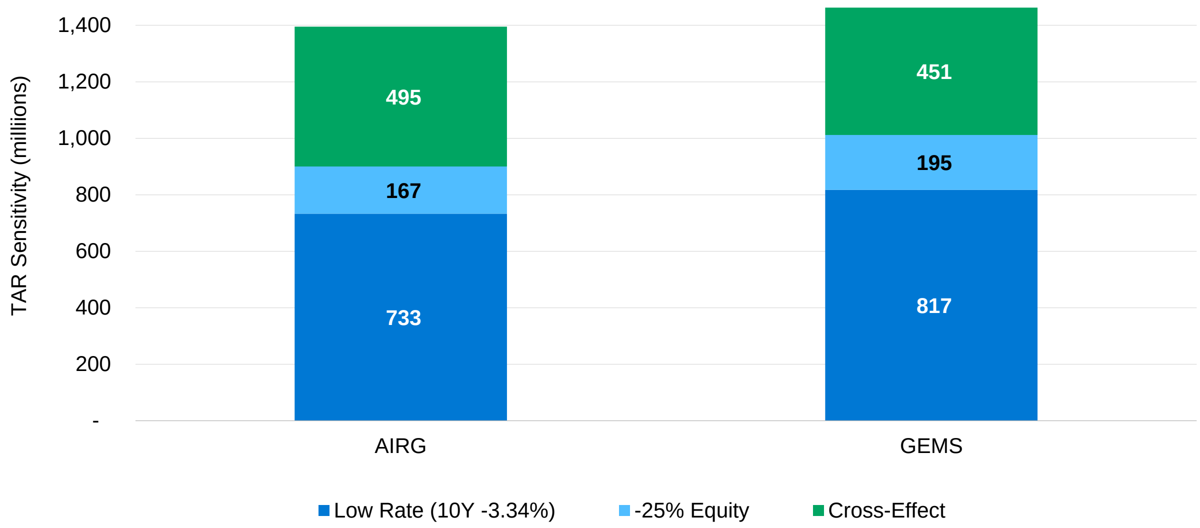
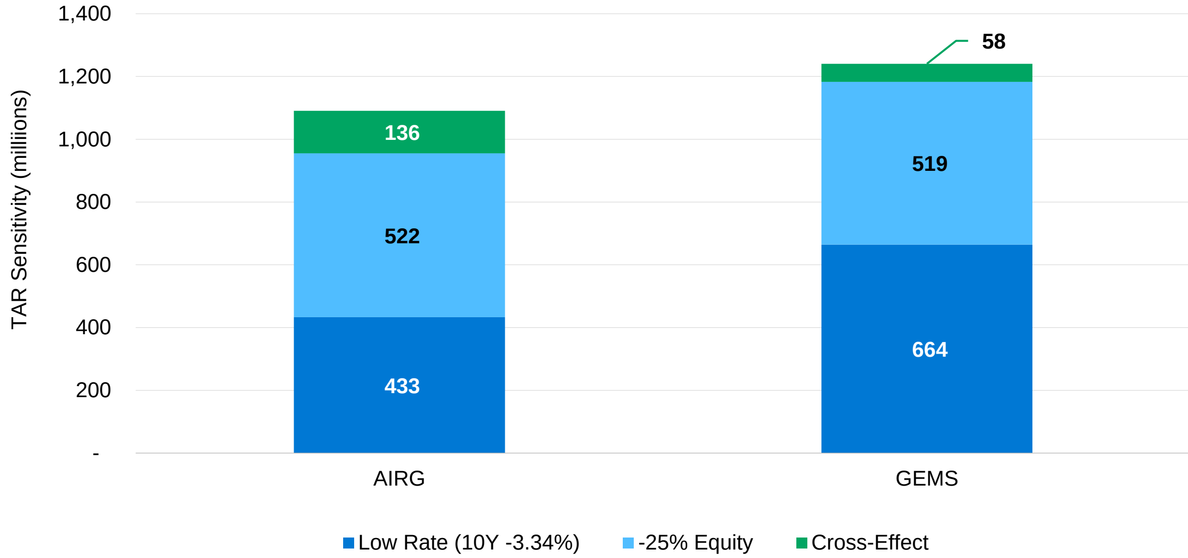


FIGURE 8: ADVERSE SHOCK TAR SENSITIVITY: UNHEDGED CTE98 TAR



The second investigation using market shocks looks at unfloored CTE70 sensitivities to measure implied "Greeks" using each scenario set. By using unfloored CTE results, we can remove distortion from CSV flooring that could lead to more convexity due to minimal sensitivity in favorable shocks. Although, in practice, companies may need to grapple with the impact of CSV flooring on their sensitivity profiles, which often can lead to a more expensive (and complicated) hedge program to operate.

Figure 9 presents tables of equity delta, equity gamma, and rho for each scenario set, hedged and unhedged. The equity Greeks were calculated based on two-sided 5% shocks, whereas the rho was unitized to a 1 basis point move and calculated based on the change in 10-year rate between the low-rate ("Field Test 2") starting curve relative to the December 31, 2023, base curve. Highlights from this table of Greeks include:

- On an unhedged basis, the delta is close between the two sets, but the results using GEMS scenarios exhibit significantly more convexity and rate sensitivity.
- On a hedged basis, all the Greeks converge relative to the unhedged basis, with AIRG and GEMS Greeks being within 3% of each other.
- Hedged CTE70 equity Greeks are lower than unhedged, whereas the rho is significantly higher. This is due to the "replacing" of real-world rider cash flows with a risk-neutral value when using an implicit hedge, particularly for the increase in rho, where the mean-reversion present in real-world GOES models dampens the long-term impact of rate shocks.

FIGURE 9: EQUITY DELTA, EQUITY GAMMA, AND RHO, BY SCENARIO SET

UNHEDGED	AIRG	GEMS	% DIFFERENCE
1% Delta	(17.54)	(18.11)	3.2%
Equity Gamma	0.22	0.25	13.7%
1bp Rho	(0.76)	(0.98)	28.5%
HEDGED	**AIRG**	**GEMS**	**% DIFFERENCE**
1% Delta	(15.03)	(14.74)	-1.9%
Equity Gamma	0.17	0.17	-1.2%
1bp Rho	(3.26)	(3.36)	2.8%

Future hedge strategy methodology

As established throughout this paper, an effectively modeled hedge strategy can blunt the impact of the GEMS scenarios, converging all scenario-level GPVADs toward the modeled hedge target. Consider the extreme case of the risk-neutral perfect hedge—intended to remove all market risk sensitivity to all liability cash flows, including equity, rate, volatility, and spread risks. In this case, every scenario would be replaced with the risk-neutral value, insulating the TAR from any market risk exposure. This is not a typical reality, as company hedge targets are often not as comprehensive and are subject to un-hedgeable risks (such as policyholder behavior) and other deviations that should be accounted for in a prudent PBR framework.

Our base case of a rider-only hedge exists along the continuum between unhedged and the perfect hedge. It assumes 95% hedge effectiveness in the best-efforts hedge, as well as a 10% E-Factor, which allows the unhedged liability cash flow impact of the GEMS scenarios to materialize. There is clear heterogeneity in the choice of hedge target in the industry, so in this section we discuss two alternative views to consider.

VEGA HEDGING WITH AN IMPLICIT HEDGE APPROACH

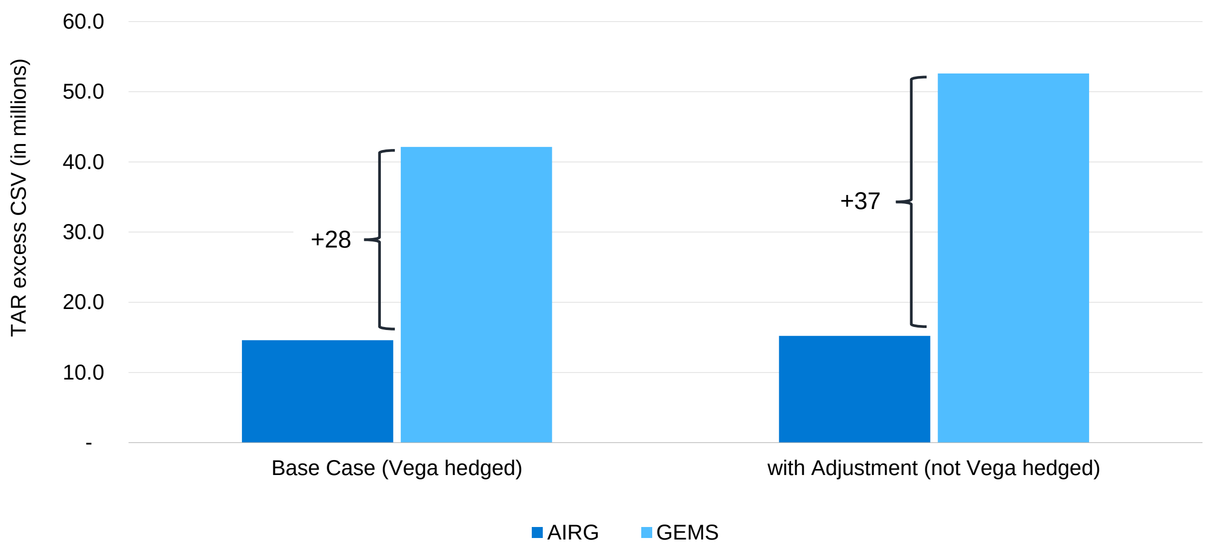
VM-21 outlines how companies must demonstrate their approach to vega hedging,[10] if vega is a hedged risk within their hedge strategy. Specifically, companies must provide evidence that the vega coverage ratio must be near 100% to justify an implicit hedge cost that uses implied volatilities, otherwise the best-efforts CTE needs to be calculated "in a manner consistent with the realized volatility of the scenarios captured in the CTE (best efforts)." Our base case assumes vega hedging is fully supported, but that may not be true for all companies. To address this, we modified the implicit hedge method to reflect a cost of hedging that varies by scenario, depending on that scenario's respective 10-year realized volatility.

10. Here, vega is the change in liability value due to changes in volatility. VM-21 Section 9.C.6 contains the details on treatment of vega hedging.

With this modified approach, the inherent difference in volatility between AIRG and GEMS scenarios flows through in the best-efforts CTE calculation, whereas it was previously neutralized. With a higher range of portfolio volatility, the hedged CTE using GEMS scenarios is penalized more by this adjustment to the hedge cost in its high-volatility scenarios.

In Figure 10, we can see that the gap between AIRG and GEMS widens with the volatility adjustment in the hedge cost. Notably, the reserves decrease with the volatility adjustment, suggesting CTE70 realized volatility is lower than the assumed implied volatilities in the base case, but the CTE98 TAR increases. This aligns with intuition, where the deepest tail scenarios are more adverse and typically experience higher volatility. This suggests that companies demonstrating vega hedge coverage can further insulate against the impact of the new GOES, but it could impact their reserve and capital allocation.

FIGURE 10: VEGA HEDGING IMPACT ON CTE98 TAR



BASE CONTRACT HEDGING

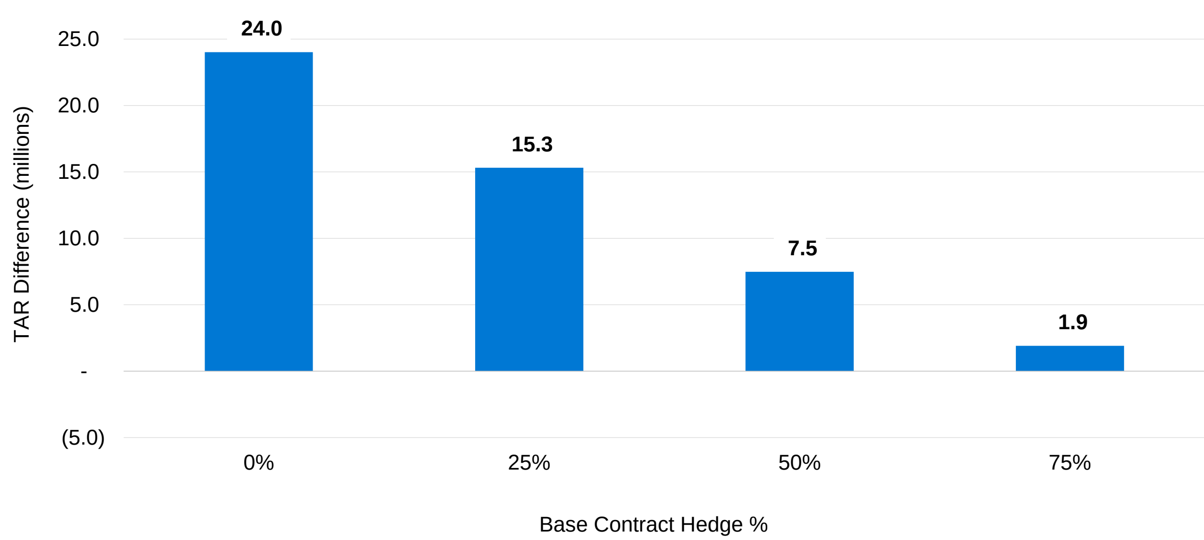
It is common practice for the VA modeled hedge strategy to only cover the rider cash flows, but as more companies in the industry have moved toward a hedging framework focused on the stability of their statutory balance sheet, some have also explored a future hedging strategy that includes the base contract cash flows. Even when rider cash flows are fully hedged in VM-21, unhedged base contract flows can exhibit significant market sensitivity, weighing adversely on tail CTE levels. To illustrate this, we introduce an alternative hedging methodology that modifies the hedge cost to include a portion of the risk-neutral value of the base contract cash flows previously unhedged in the base case.

In Figure 11, we present the impact on the unfloored best-efforts CTE70 between the scenario sets from hedging different percentages of the base contract cash flows. Unfloored CTE results are used because the floored results are primarily floored at the CSV, muting the effect of the base contract. This can provide a more illustrative view of the impact of hedging varying proportions of the base contract.

As the hedge ratio on the base contract increases, TAR under the AIRG and GEMS scenarios converges. This tracks intuition, as a larger share of the total liability cash flows are being "replaced" by the risk-neutral value (rather than passed through based on the realized scenario), less variation remains between the liability in each scenario. Some variation remains due to assumed hedge ineffectiveness. This illustrates another way in which companies' hedge strategies can insulate them from potentially significant scenario changes—the more closely they hedge the full contract to a risk-neutral basis, the less exposure to the underlying scenario set.

This does come at a cost, though. Even though the TAR impact between scenario sets is much smaller, the starting TAR level is impacted. For instance, on the AIRG scenario set, hedging 100% of the base contract leads to a $69 million increase in unfloored CTE70 relative to only hedging the rider. Further, a hedge strategy that fully hedges base contract cash flows may strain profit forecasts along deterministic paths with assumed favorable equity returns.

FIGURE 11: GEMS-AIRG DIFFERENCE: UNFLOORED BEST-EFFORTS CTE70



Conclusion

Given the complexity of VA business and the diversity of risk factors across companies, no model is going to perfectly capture the impact of the proposed GOES scenarios on statutory reserves and capital. However, the impact analysis presented in this paper intends to shine light on common themes that should be representative of what companies might expect. Those themes include:

- An increase in reserves and capital, which could be significantly reduced if an effective future hedge strategy is reflected in the CTE calculation.
- Similar equity sensitivity, but increased rate sensitivity with GEMS scenarios.
- Less exposure to the GEMS scenario impact if the future hedge strategy supports a full vega hedge or includes base contract cash flows in the hedge target.

As the GEMS scenarios are finalized, companies should begin evaluating the implications of the new scenarios on not only their required levels of reserves and capital, but also on other key metrics and risk management practices. For example, the change in market shock sensitivities could necessitate an adjusted hedge portfolio. Or companies may need to consider the impact of how fund exposure maps onto the new array of equity and fixed income funds offered on the GEMS model. Projected reserves and capital could be materially affected, too, and companies could evaluate this with a stylized “constant rates” path that forecasts pro forma earnings with reserves and capital based on the field test scenarios. Through detailed analyses like these, companies will gradually become prepared for the new GOES model once the NAIC gives final word on implementation parameterization and timing. Doing so will enable them to develop the technical and business prowess necessary for successful risk management going forward.

Solutions for a world at risk™

Milliman leverages deep expertise, actuarial rigor, and advanced technology to develop solutions for a world at risk. We help clients in the public and private sectors navigate urgent, complex challenges—from extreme weather and market volatility to financial insecurity and rising health costs—so they can meet their business, financial, and social objectives. Our solutions encompass insurance, financial services, healthcare, life sciences, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

milliman.com

CONTACT

Collin Davidson
collin.davidson@milliman.com

Parker Henley
parker.henley@milliman.com

Nick Lee
nick.lee@milliman.com

Zohair Motiwalla
zohair.motiwalla@milliman.com