A Darwinian Theory of Model Risk

Claudio Albanese[1], Stéphane Crépey[2], and Stefano Iabichino[3]

May 27, 2021

Abstract

Performance assessment of derivative pricing models revolves around a comparative model-risk analysis. From among the plethora of econometrically unrealistic models, the ones that survive the Darwinian selection tend to generate systematic short-term profits while exposing the bank to long-term risks.

This article proposes an ex-ante methodology to analyze the model-risk pattern for the broad class of structures, whereby a dealer buys long-term convexity from investors and resells hedges for risk management purposes. As a particular case, we consider callable range accruals in the US dollar, a product that has been traded in size in recent years and is currently generating material losses. To visualize the sources of model-risks, we use 3d animations.

1 Principles

The objective of derivative portfolios' pricing and hedging drives models' development. The risk in a derivative portfolio can be partitioned between market and model risk. The former focuses on the second or higher moments of the return distribution and is often managed by overnight Delta or Vega hedging strategies. Contrary, as we will discuss, model-risk revolves around the first moment and results in systemic and autocorrelated leakages that manifest as a bias over longer periods.

If model-risk is only analyzed on a backward-looking basis, a realized leakage is interpreted as a "black swan" event, an interpretation that precludes disentangling sound and unsound models. Along this line of thought, a trader always perceives upfront profits as realized, and model blow-ups take the role of rare and unlucky events he has no margin of maneuver to prevent.

[1] Global Valuation, London.

[2] LPSM, Université de Paris, Building Sophie Germain, Avenue de France, 75013 Paris. The research of Stéphane Crépey has been conducted with the support of a grant Derivative Accounting Principles at the Test of the Covid-19, under the aegis of the Europlace Institute of Finance, France.

[3] JP Morgan. This paper represents the opinions of the authors and it is not meant to represent the position or opinions of JP Morgan or their members.

Acknowledgement: We thank two anonymous referees and Christian Kappen for their useful comments.

This article discusses the Darwinian theory of model-risk, i.e., the antipodal opposite of the black swan theory, providing a scientific explanation of financial blow-ups based on models' behavioral analysis. In more detail, we develop a forward-looking model-risk framework for structured products sharing similar features. By trading structural products, a bank buys volatility and convexity from investors and sells out-of-the-money (OTM) options trading at a premium for risk management purposes. Model-risk manifests in yield enhancement features, such as optional callability or auto-callability, without which the bank could statically replicate these products without incurring any model-risk. Examples include callable range accruals hedged by digital swaptions (our case study), structured equity products such as autocallables and cliquets, power-reversal dual currency options, and target redemption forwards.

The present article uses a forward-looking state-space analysis to discover models blow-up patterns. Instead, in Albanese, Crépey, and Iabichino (2020) we precisely identify blow-up patterns via a reverse-engineering exercise, and we extend our analysis to the XVA metrics generated by a portfolio of exotic derivatives.

Quality ranking for models is subjective and a function of the utility of the model user. Typically, a user with interests aligned with the bank would prefer more realistic models, which have lower short-term profitability but far less long-term risk. However, a bad actor might prefer a lower-quality unrealistic model as its usage could help extracting more short-term wealth in exchange for long-term risk.

From among the vast plethora of econometrically unrealistic models, the ones that stand a chance to survive and become broadly used have to satisfy two adverse selection conditions. The first condition relates to competitive pricing:

First Darwinian Principle *A lower-quality model surviving the test of time must over-value the structured product at inception.*

An over-valuation at inception implies that the upfront payment paid to the investor could be greater than the payment implied by a more sound model. The over-valuation makes the model competitive in the market, a necessary model's survival condition, or else the trader would discard the model as its usage would attract no business.

However, in line with the Doob-Meyer theorem decomposition of supermartingales, the mirroring companion of over-valuation at inception is a negative alpha-leakage as time goes on. As reality unwinds and the model is re-calibrated, the initial over-pricing turns in valuations' systematic decay, making realized discounted price processes deviate from martingales, i.e., a signal of the existence of arbitrage opportunities on an ex-post basis. Hence, the aggressive initial valuation must turn into a mark-to-market loss for the user as reality unavoidably sets-in.

The first Darwinian principle raises a profitability puzzle: how can the systematic PnL losses be sustained over time? Over-valuation is a survival's necessary, yet not sufficient, condition. We need a second condition to guarantee the survival of the "fittest".

Second Darwinian Principle *A surviving lower-quality model will over-hedge and,* *limited to the short to medium term, the profits generated by hedge positions will overall offset and surpass the systematic losses engendered by the initial over-valuation.*

The second Darwinian principle conceals the first Darwinian principle's losses, at least in the short to medium run. If a model satisfies the Darwinian principles, it stands a chance at survival and might become an accepted derivatives market pricing standard.

Alpha leakages are undetectable by market risk tools such as VaR, Expected Shortfall, and Stressed VaR models as these frameworks focus on the return distributions' second or higher moments over short-time horizons. Backward-looking statistical and Machine Learning methods estimate only the realized alpha terms, and, in the worst case, they take the role of post-mortem forensic tools. Contrarily, the forward-looking state-space analysis of this paper anticipates the risk, as model-risk losses surface using a Challenger model to simulate the Champion model's hedging strategy.

2 Equations

In what follows $\mathcal{M}$ denotes a continuous martingale that changes from line to line. We develop a stylized setup where trading book valuations follow a process P such that

$$dP_t = r_t P_t dt - \alpha_t^P dt + d\mathcal{M}_t, \tag{1}$$

where r is the risk-free rate and $(-\alpha^P) \leq 0$ is a valuation alpha leakage. A simple framework for model risk involves looking at the valuation process from the dual viewpoint of a sound model (G) and an unsound model (B). At time 0, both models are assumed to calibrate well to the same set of instruments used for hedging. Equation (1) is written under the sound model's measure and expresses the valuation's dynamics process P computed using the unsound model[1]. The good model calibrates well to the accounting exit price (fair value as per IFRS 13) of the exotic derivative at all times.

Models have a dual purpose: pricing for the client and generating a hedging strategy. The valuation provided by the bad model is typically above the exit price, except that the difference is placed in a model risk reserve account and is not recognized in PnL. Structured products, such as the callable range accrual in our case study, can be profitable to the bank because their hedging strategy involves selling out-of-the-money options to buyers whom we know are willing to pay a premium. As a consequence one would expect that the hedging strategy executed with a perfect model would still produce systematic gains.

Nevertheless, a bad actor could seek to generate additional gains in the short term by playing with model switches (see Figure 1). Namely, the bad actor could use a sound

[1]The case $\alpha^P = 0$ corresponds to the sound model case

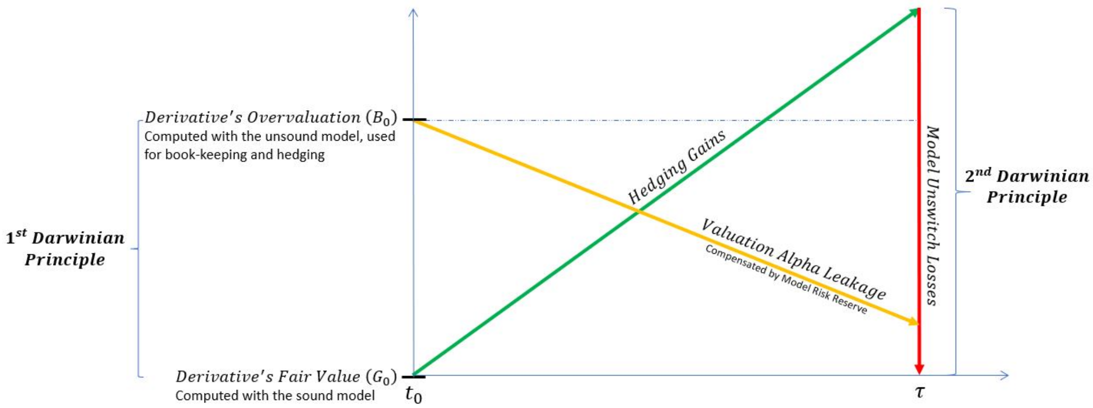

Figure 1: Model risk stylized pattern (in red PnL losses, in green PnL profits, in orange model risk reserve)

model, corresponding to $\alpha^P = \alpha^G = 0$ in (1), for fair-value discovery, and a unsound model, with some $\alpha^P = \alpha^B > 0$, to determine both the entry-price upper-bound and the potential over-hedging gains. Therefore, model B would be used for the sake of increasing volumes, hedging and book-keeping.

However, a risk manager would insist that the difference $B_0 - G_0$ is retained in a model risk reserve to avoid monetizing a fictitious instant PnL gain to the trader's benefit (generated by the book-keeping usage of the B model). Ultimately, the strategy's valuation side will reduce to a negative valuation drift $(-\alpha^B)dt$ in the bank PnL, which the model risk's reserve absorbs.

But the key point of this paper is that the model-risk reserve strategy is often insufficient. The unsound model does not only influence valuation (cf. the **First Darwinian Principle**); it also affects hedging. Typically, unsound models generate over-hedging (cf. the **Second Darwinian Principle**), which generate material short-term profits if they expire worthless, while giving rise to significant losses if they don't.

To cut losses in case excessive hedges appreciate, a risk manager would typically place trading limits on the over-valuation amount and/or on the aggregate value of the hedges. In the event that trading limits are hit, the bank will (i) unwind the hedges, (ii) recognize a portfolio mark down and (iii) switch away from the unsound model, at a substantial loss. This tends to happen at time of market stress when the hedges are rapidly appreciating and the rush to buy them back from other dealers can trigger a nonlinear blow up by a compounding effect, i.e. a so called "gamma trap".

Suppose $\mathcal{H}$ denotes the cumulative value process of excessive hedges generated using the unsound model. The cumulative wealth in the hedging account satisfies an equation of the form $\mathcal{H}_0 = 0$ and

$$d\mathcal{H}_t = r_t \mathcal{H}_t dt + (\alpha_t^L + \alpha_t^I)dt + dJ_t^- + d\mathcal{M}_t, \tag{2}$$

where $\alpha^L, \alpha^I \geq 0$. The term α^L is a legit profit, as driven by the actual derivative's hedging requirements. The term α^I is illegit,"Ponzi profit", as it simply acts as a compensator for the downward jump process J^-, i.e., the process

$$\alpha_t^I dt + dJ_t^- \tag{3}$$

is a (risk-neutral) martingale. The loss dJ_t^- materializes at the time τ where the unsound model stops being viable, as trading limits are triggered because the overvaluation implied by the bad model over the exit (good model) price is excessive. Note that the ensuing loss, and therefore α^I in the above, is driven by α^B.

The overall PnL of the bank, V say, satisfies $V_0 = 0$ and, for $t \geq 0$,

$$dV_t = r_t V_t dt + (\alpha_t^L + \alpha_t^I - \alpha_t^B)dt + dJ_t^- + d\mathcal{M}_t, \tag{4}$$

where, by the **Second Darwinian Principle**,

$$\alpha_t^L + \alpha_t^I - \alpha_t^B > 0. \tag{5}$$

The accumulated amounts $(\alpha_t^L + \alpha_t^I - \alpha_t^B)dt$ would also drive the trader's remuneration until time τ. In the short to medium term, the bank might not perceive the risk represented by the term J^- in V and see the strategy as globally profitable in view of (5). But the actual drift of the strategy is only $(\alpha_t^L - \alpha_t^B)dt$ and a loss dJ_t^- materializes. at the time where the unsound model is not sustainable and the bank has to unwind hedges and mark-down the portfolio. At this time the trader takes no part to the jump loss, even though he diverted a part of the jump compensator as his benefit.

By comparison, without the model switch strategy, the PnL of the bank, $\tilde{V}$ say, would satisfy $\tilde{V}_0 = 0$ and, for $t \geq 0$,

$$d\tilde{V}_t = r_t \tilde{V}_t dt + \tilde{\alpha}_t^L dt + d\mathcal{M}_t. \tag{6}$$

hence no downward jump risk and a purely legit drift coefficient[2] $\tilde{\alpha}^L$.

One may even argue that, if unsound models and model switches are allowed, our "bad actor" might not be so much to blame: he could have no choice than using a unsound model to ensure his survival as he would operate in a competitive equilibrium generated by the usage of unsound models. By contrast, if only sound models are allowed for pricing, hedging, and book-keeping purposes (indeed, the practice we advocate) or model switches are forbidden, the competitive equilibrium will revolve around sound models, immunized against systemic PnL blow-ups. Hence, only an external regulator can break the unsound models usage deadlock to the mutual benefit of the bank's traders, shareholders, and banking stability.

3 The Case of Callable Range Accruals

Hedging strategies typically consist of the combination of a robust strategy, which is model-independent (see for instance Carr and Madan (1998), Dupire (1993), Hagan

[2]Typically, $\tilde{\alpha}^L < \alpha^L$ in (4) as a bad actor would also leverage on unsound models to enhance liquidity gains via distorted hedging ratios.

(2002)), plus a model-driven strategy which is subject to model risk. Examples of model risk sources are optionality clauses embedded in derivative contracts such as early termination as in American style options, where different models typically lead to different conclusions regarding optimal termination decisions. Typically, different models would lead to different conclusions regarding optimal termination decisions and risk sensitivities.

In this section, we illustrate the case of a callable range accrual under the Darwinian principles's spotlight. A range accrual is a derivative product very popular among structured-note investors. The investor in a range accrual bets that the reference underlier, usually interest rates, will stay within a predefined corridor. Callability is commonly added to enhance the returns and make the product more attractive to the client, but this is also the place in which Darwinian model risk unveils.

3.1 The Payoff

Consider a range accrual issued at time 0 with maturity T. A typical payoff at time $t \in [0, T]$ is $\Phi_t dt$ to the bank, where

$$\Phi_t = K_0(I_s^- + I_s^+) - K_1.$$

Here $I_s^- = \mathbb{1}_{\{I_s < R_0\}}$ and $I_s^+ = \mathbb{1}_{\{I_s > R_1\}}$, where $[R_0, R_1]$ is a range and I_s is an index process, typically of the form

$$I_s = \eta_1 \mathrm{SR}_s^{(1)} - \eta_2 \mathrm{SR}_s^{(2)},$$

where $\mathrm{SR}_s^{(1)}$, $\mathrm{SR}_s^{(2)}$ are two swap rates and $0 \leq \eta_1, \eta_2 \leq 1$ are constants. The swap rates are functions of the time s discounting curve and Libor rates. Hence $I_s^{\pm}$ are payoffs of European style digital swaptions in the index maturing at time s.

Some range accruals are callable by the issuer as, in order to entice the investor with a larger coupon, the bank reserves the right to terminate the transaction prior to maturity. From the investor viewpoint, termination could be acceptable since it will occur only in a state of the world where the bank has incurred in large losses, hence the investor has received excess revenues.

The time-0 valuation of a callable range accrual under the risk neutral pricing measure, with corresponding expectation $\mathbb{E}$, is given by (we skip the risk-neutral discount factors for notational simplicity)

$$\Pi_0 = \mathbb{E}\big[\int_0^T \mathbb{1}_{\{s \leq \vartheta\}} \Phi_s ds \big],$$

where ϑ is the stopping time for the early exercise held by the bank.

Non callable range accruals are relatively simple to analyze as the issuer can statically hedge them by writing a set of digital swaptions struck at the corridor boundaries (see Figure 2).

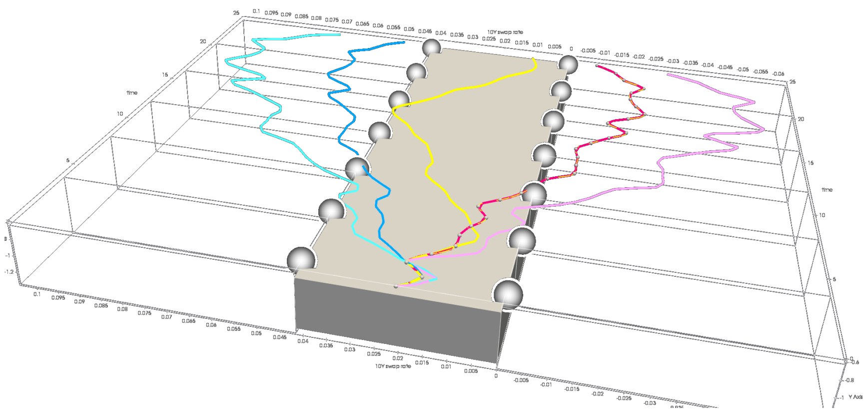

Figure 2: Range accrual corridor (the grey box), static hedges (on the sides of the corridor), and sample paths.

In this case, we have that $\Phi_s = K_0(I_s^- + I_s^+) - K_1$. Hence the valuation at the inception of the non-callable variant of the range accrual can be decomposed as follows

$$P_0 = \int_0^T \mathbb{E}(\Phi_s)ds = \int_0^T \big(K_0(D^+(s) + D^-(s)) - K_1\big)ds,$$

where $D^\pm(s) = \mathbb{E}(I_s^\pm)$.

The time 0 valuation of a callable range accrual instead is given by

$$\Pi_0 = \int_0^T \mathbb{E}\big[\mathbb{1}_{\{s<\vartheta\}}\big(K_0(I_s^+ + I_s^-) - K_1\big)\big]ds = \int_0^T \big(K_0(\Delta^+(s) + \Delta^-(s)) - K_1\Xi(s)\big)ds.$$

Here $\Delta^\pm(s) = \mathbb{E}[\mathbb{1}_{\{s<\vartheta\}} I_s^\pm]$ are time-0 valuations of "Canadian style" digital swaptions (cf. Carr (1998)), which expire worthless in case the callable range accrual is exercised, and $\Xi(s) = \mathbb{E}\big[I_{s<\vartheta}^-\big]$. Assuming that digital swaptions are fairly priced, a quasi-robust hedging strategy for the callable range accrual in terms of European digital swaptions is given by the following static hedging ratios (constant over time t) in the digital swaptions with infinitesimal ds nominals (and the ensuing funding position so as to make the strategy self-financing):

$$a^\pm(s) = \frac{\Delta^\pm(s)}{D^\pm(s)}.$$

We say quasi-robust, because the time-t hedging ratios that would correspond to the same analysis as above, but starting from a time t in the future, are quite stable through time t. In practice, people use these hedging ratios plus a classical Delta hedging strategy for the residual hedging error already "smoothed out" by the digital swaptions quasi-static hedge.

3.2 Champion and Challenger

For the model-risk analysis, we compare two models: a champion and a challenger.

The champion is the Hull-White (HW) 1-factor model specified as

$$dr_t = \kappa(t)(\theta(t) - r_t)dt + \sigma(t)dW_t.$$

In the HW model, rates are not bounded from below and the classic Hull and White (1990) solver depends on this assumption.

The challenger is a model with two fundamental differences: rates are bounded from below and there are two factors instead of one (so that the steepness of the yield curve is stochastic). An example of such model is the stochastic drift (SD) short rate model of Albanese and Trovato (2008), where the short rate process is defined by an equation of the form

$$r_t = \phi(t) + \lambda(t)\rho_t. \tag{7}$$

Here

$$\phi(t) = \min(0, f(t) - 20\text{bp}), \tag{8}$$

where $f(t)$ is the infinitesimal forward rate, $\lambda(t)$ is a drift adjustment factor, and

$$d\rho_t = \kappa(\theta_t - \rho_t)dt + \sigma(t)\rho_t^{\beta} dW^{(1)}, d\theta_t = k(a - \theta_t)dt + \nu dW^{(2)}, \tag{9}$$

with $d\langle W^{(1)}, W^{(2)}\rangle_t = \gamma dt$. The resulting short rate model is very similar to a shifted 2-factor Libor market model.

To make the qualitative comparison significant, we select specific instances, sharing the properties of the HW and the SD model, of a common framework of discrete hidden Markov models (see Albanese and Trovato (2008)).

Figure 3 reports the differences in the calibration accuracy between the SD and the HW models for the 5-year swaption surface (in bps). We refer to Albanese and Trovato (2008) for details on the models' calibration routine. Figure 4 depicts the calibrated models' probability densities as a function of time, from which the major differences between the two models can be observed.

	-200	-150	-100	-75	-50	-25	0	25	50	75	100	150	200	300	1000
31		-81.25	-55.39	-37.76	-19.65	-7.78	-8.66	-12.63	-23.54	-38.03	-53.40	-85.79			
92	-41.27	-30.63	-16.63	-9.69	-5.30	-3.70	-7.32	-8.77	-11.48	-15.33	-20.26	-32.00	-44.48	-70.15	
184	-16.86	-10.83	-4.01	-1.68	-0.57	-0.57	-3.32	-4.46	-6.14	-8.24	-10.57	-16.44	-23.12	-36.62	
365	-3.59	-1.26	-3.59	1.30	1.37	1.07	-1.03	-1.87	-2.94	-4.23	-5.72	-9.16	-12.89	-20.60	-75.22
731	3.81	3.40	3.01	2.71	2.37	1.95	0.11	-0.46	-1.14	-1.95	-2.86	-4.84	-6.94	-11.33	-44.59
1096	5.07	3.81	2.78	2.29	1.83	1.37	-0.57	-1.07	-1.72	-2.40	-3.09	-4.50	-5.91	-8.81	-32.16
1460	4.54	2.90	1.53	0.95	0.38	-0.19	-2.14	-2.71	-3.32	-3.93	-4.50	-5.57	-6.56	-8.51	-24.11
1824	5.91	4.01	2.48	1.75	1.07	0.42	-1.60	-2.21	-2.78	-3.28	-3.78	-4.58	-5.30	-6.60	-16.94
2557	5.49	3.55	1.91	1.14	0.46	-0.15	-1.79	-2.33	-2.78	-3.20	-3.51	-4.08	-4.50	-5.04	-8.70
3651	11.10	9.16	7.59	6.90	6.26	5.72	5.26	4.92	4.65	4.42	4.31	4.12	4.08	4.27	6.10
5478	11.33	9.61	8.28	7.71	7.17	6.71	7.13	6.87	6.71	6.60	6.56	6.48	6.52	6.87	9.80
7305	8.96	7.55	6.48	6.03	5.61	5.23	5.49	5.30	5.19	5.11	5.07	5.00	5.04	5.19	6.60
9131	10.53	9.31	8.28	7.82	7.48	7.13	7.44	7.32	7.29	7.36	7.40	7.59	7.86	8.32	10.91

Figure 3: HW model calibration accuracy in bps observed from the SD model: 5-year swaption surface (columns: Swaptions' tenor; rows: Swaptions' moneyness)

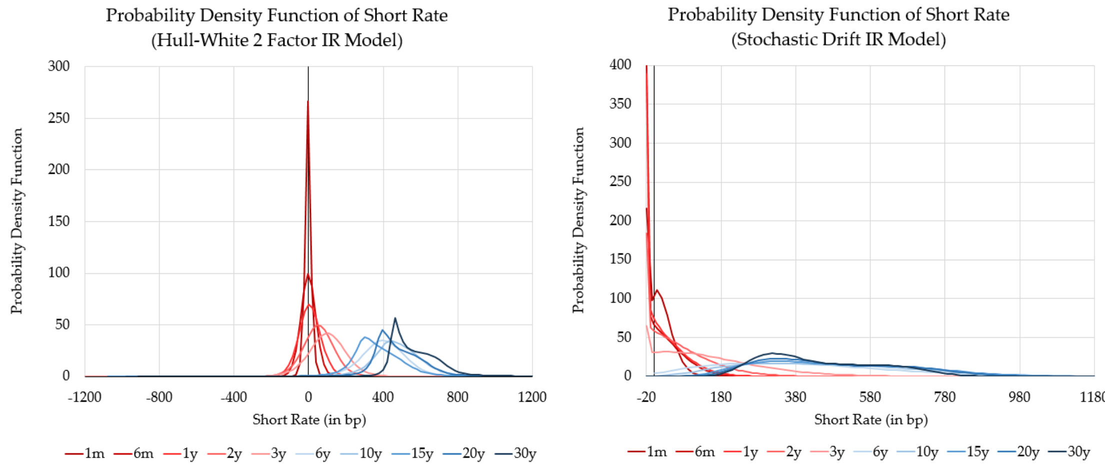

Figure 4: Probability densities of the USD short rate computed using the HW (left) and SD (right) models. Probability densities of the USD short rate computed using the champion (left) HW model and challenger (right) SD model.

3.3 Alpha-leakages in action

We run the SD and the HW models on a callable range accrual in the USD of maturity 25 years and with corridor $[0, 4\%]$. 3d animations are found the best way to visualize sources of model-risk, and we refer to this sample model risk report, see Albanese (2019). The frames corresponding to months 10, 100 and 200 have the following interpretation:

- The comparison between the challenger's (below) and the champion's (above) valuations, which differences represent the valuation alpha leakages, as a function of the future possible states of the economy (see Figure 5)
- The valuations of the challenger model are then investigated in terms of the cross gamma eigenvector (see Figure 10);
- The HW model's and the SD model's optimal exercise boundary (see Figures 8 and 6 respectively);
- The comparison between the champion's and the challenger's hedge ratios (see Figure 7). The champion model's overhedging requirements are visualized from the challenger model's viewpoint (see Figure 11);
- The challenger's least crossgamma eigenvalue (see Figure 9).

The ensuing comparison between the two models are summarized in Table 1.

The HW model's alpha-leakages materially dominate the SD model's. The alpha-leakage dominance relates to the profound negative rates, which is probable in the HW model while not possible in the SD model. When rates descend to deep negative values in the HW model, range accruals turn into a precious asset for the issuer. The rates'

Gaussian assumption raises the valuation of the HW model also in the positive range (see Figure 5).

	HW Model	SD Model
Valuations (cf. the first Darwinian principle)	The unbounded short-rate assumption has a direct impact on the HW's alpha leakages. Alpha leakages are particularly pronounced in states characterized by deep negative rates, in which the range accrual is predicted as particularly valuable for the issuer.	The bounded short-rate process in the SD model constraints the alpha-leakages, especially in scenarios characterized by low-interest rates. In low-rates scenarios, the probability of exiting the corridor, which would turn the range accrual into an asset for the issuer, is immaterial, and there is no economic value in keeping the position open.
Early exercise	Low-interest rates scenarios are of great value for the HW model. In low-interest rates scenarios, the optimal exercise decision is postponed as the probability of having deep negative interest rates, in which the range accrual would turn into an asset for the bank, is material. Therefore, there is no economic value in early unwinding the trade for the issuer.	The SD model perceives low-interest rates scenarios as the riskiest states, as the probability to exit the range accrual's corridor is immaterial. Therefore, the SD model signals the optimality in early unwinding the trade. The optimal exercise boundary in the SD model is greatly affected by the steepness of the interest rate curve (see Figure 6), a feature which is not captured by the HW model.
Hedge ratios (cf. the second Darwinian principle)	HW's higher hedging ratios are related to the delay in which the optimal exercise decisions are signaled by the HW model.	SD model's hedging ratios, in the digital swaptions struck at the corridor boundaries, are far smaller than the hedge ratios implied by the HW model (see Figure 7).
Sensitivities	For rates within the corridor, the HW model shows mostly positive gammas, signaling that dynamic hedging is relatively safe.	For rates within the corridor, the SD model indicates that the cross-gamma matrix has mostly negative eigenvalues (see Figure 9), signaling the presence of gamma traps. Once represented in terms of the 10y and the 10y-2y swap rate, the lowest eigenvalues' angle varies from 30°to 70°as the corridor boundary is explored (see Figure 10).

Table 1: Champion / Challenger comparison.

In particular, the HW model depicts a somehow appealing story from the bank's viewpoint, especially in low-interest-rate regimes. The gamma is positive across most of the state-space, and the option termination occurs significantly late in the lifetime, allowing the bank to seize the benefit of carrying of the hedges, which are traded at a material liquidity premium. However, from the viewpoint of the SD model, the excessive hedges represent a financial risk (see Figure 11).

Over the most profitable period for the trade, the risk profile due to excessive hedges exhibits a saddle point shape centered in the middle of the corridor and a pronounced steep, which signals potential losses in case of falling rates. This situation unveils the typical cross-gamma risk, whereby one eigenvalue of the cross-gamma matrix is positive, and the other is negative (see Figure 9 and 10).

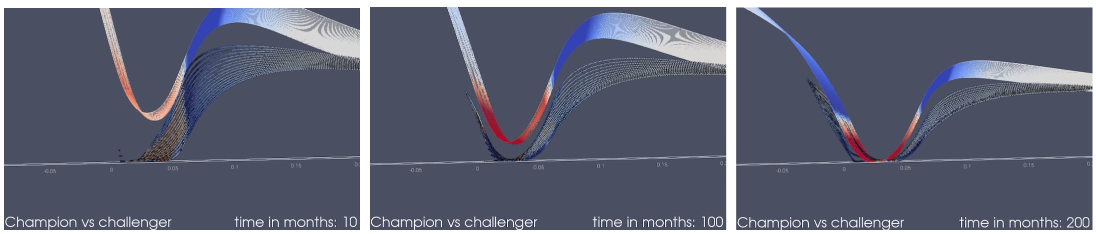

Figure 5: Comparison between the valuations of the range accrual by the challenger (below) and the champion (above), color-coded for the gamma (red for positive gamma, white for intermediate, blue for negative. X-axis represents the 10y swap rate; Y-axis the 10y-2y swap rate spread; the Z-axis is the values of the range accrual).

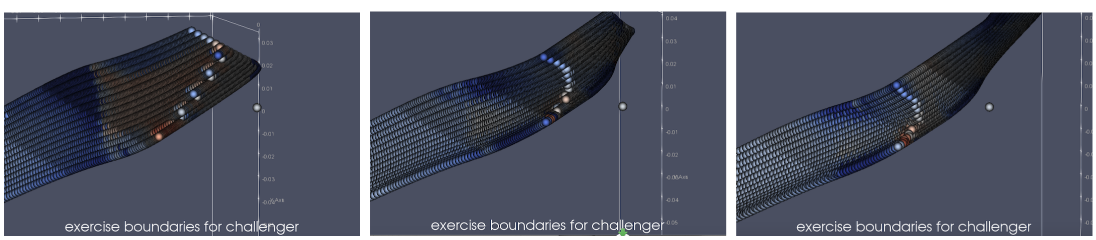

Figure 6: Exercise boundaries of the challenger. (X-axis represents the 10y swap rate; Y-axis the 10y-2y swap rate spread; the Z-axis are the range accrual's valuations).

The HW model, proposed in 1990, was a veritable breakthrough as it was the first example of a class of term structure models later generalized in Heath, Jarrow, and Morton (1992). Within the HJM models class, the HW model and its 2-factor extensions are unique in the simplicity of implementation. Three decades ago, in a limited technological environment, the HW model's solvability was the model's biggest strength, a strength achieved by an underlying Gaussian structure. Three decades later, in a technology environment where closed-form solvability is no-longer a model's binding constraint, the Gaussian hypothesis, which implies unbounded interest rates from below, turned the HW model's strength into an Achille's heel.

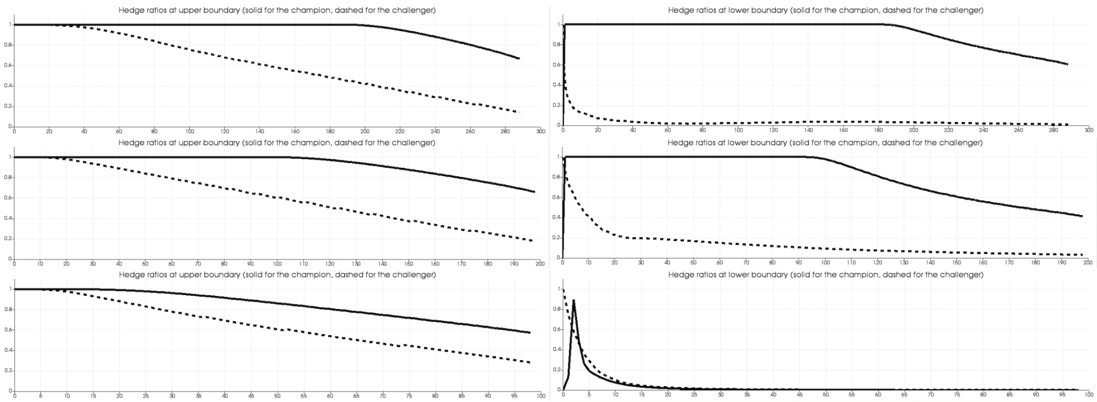

Figure 7: Comparison of the hedge ratios for the champion and the challenger.

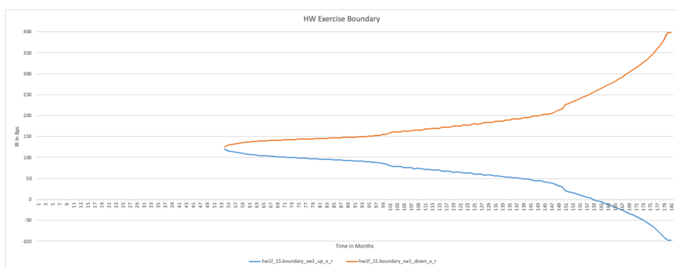

Figure 8: Exercise boundaries of the champion. (X-axis time in months to maturity; Y-axis 10y swap rate).

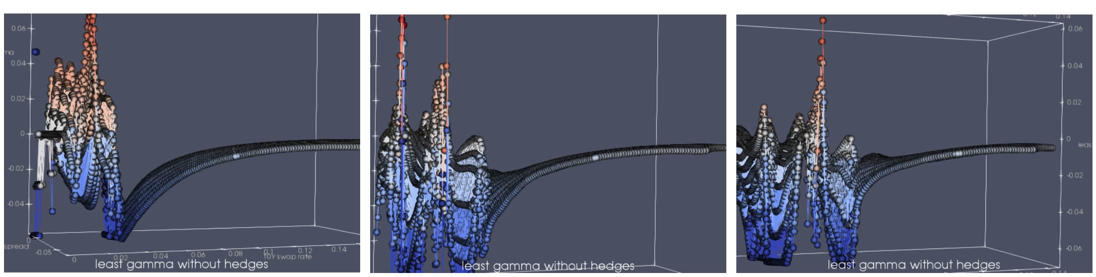

Figure 9: Least crossgamma eigenvalue in the SD model (the challenger), color-coded as blue if negative, red if positive (X-axis represents the 10y swap rate; Y-axis the 10y-2y swap rate spread; the Z-axis is the cross-gamma value).

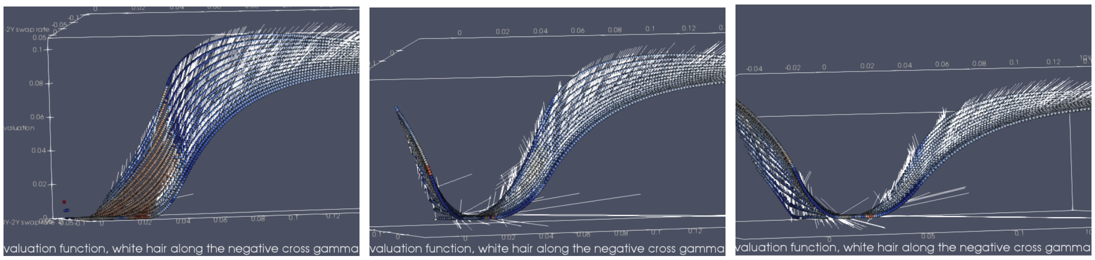

Figure 10: Valuations of the challenger. The white filaments point in the direction of the negative crossgamma eigenvector (X-axis represents the 10y swap rate; Y-axis the 10y-2y swap rate spread; the Z-axis is the values of the range accrual in the SD model).

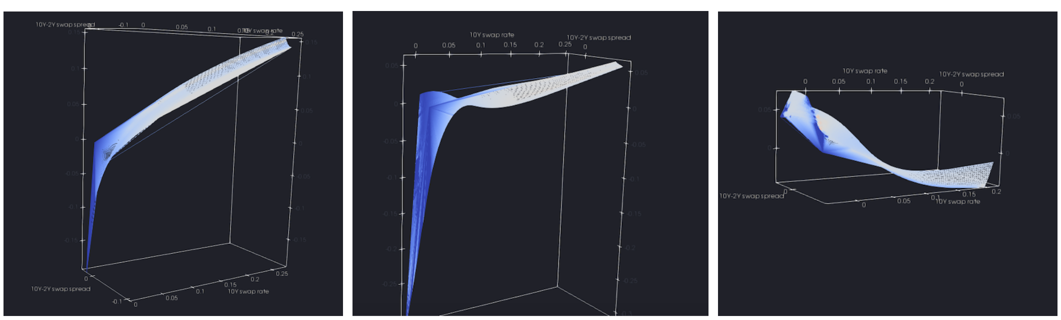

Figure 11: Exposure to the over-hedging as quantified by the valuation of the excess of short digital swaptions implied by the HW model from the viewpoint of the SD model (X-axis represents the 10y swap rate; Y-axis the 10y-2y swap rate spread; the Z-axis is the value of the portfolio of excessive hedges).

4 Conclusion

Darwinian selection favors models that generate systematic profit in the short and medium-term compensated by possible large losses in the long term. Alpha leakages are undetectable by market risk models such as Value-at-Risk (VaR), Expected Shortfall, and stressed VaR. Backward-looking statistical and Machine Learning methods estimate only the realized α terms. A forward-looking analysis anticipates also the jump; model-risk losses become apparent using a Challenger model to simulate the hedging strategy of the Champion model. The state-space analysis on which this article basis, facilitates model-risk understanding by employing intuitive visualizations and 3d animations of the model-risk's sources. We conclude that the preference for short-term profitability skews returns giving rise to large losses in stress conditions.

Model-risk is intimately intertwined with reverse stress testing, the forward-looking variant of stress testing. A model can not be considered as valid if it is known to break down on a path leading to a stress scenario. While in the present article we are concerned with understanding the pattern that leads to blow-up under stress, in Albanese, Crépey, and Iabichino (2020) we focus on the discovery of models' blow-up scenarios.

References

Albanese, C. (2019). `https://www.youtube.com/watch?v=CVVbcEfaEBU&t=331s`.

Albanese, C., S. Crépey, and S. Iabichino (2020). Reverse stress testing. Working paper, Available at SSRN: https://ssrn.com/abstract=3544866.

Albanese, C. and M. Trovato (2008). A Stochastic Monetary Policy Interest Rate Model. *Computational Methods in Financial Engineering*, 343–392.

Carr, P. (1998). Randomization and the American put. *The Review of Financial Studies 11*(3), 597–626.

Carr, P. and D. Madan (1998). Towards a theory of volatility trading. In *Volatility: new estimation techniques for pricing derivatives*, pp. 417–427. Risk Books.

Dupire, B. (1993). Model art. *Risk Magazine 6*, 118–124.

Hagan, P. (2002). Turning good prices into great prices. *Wilmott Magazine 2*, 56–59.

Heath, A., R. Jarrow, and A. Morton (1992). Bond pricing and the term structure of interest rates: A new methodology. *Econometricas 47*.

Hull, J. and A. White (1990). Pricing interest rate derivative securities. *Review of Financial Studies 4*.