Incorporating Smile in Valuation Adjustments Through the Mixture of Short-Rate Models

Utrecht University & Rabobank, the Netherlands

T. van der Zwaard

2023 SIAM Financial Mathematics and Engineering - June 7, 2023

Acknowledgements & Disclaimer

Joint work with L. Grzelak (UU, Rabobank) & C. Oosterlee (UU).

Acknowledgements
This work has been financially supported by Rabobank.

Disclaimer
The views expressed in this work are the personal views of the authors and do not necessarily reflect the views or policies of their current or past employers.

Outline

Goal: incorporate smiles in Valuation Adjustments (xVAs).

Steps:

1. Introduction.
2. Our contribution.
3. SDE with state-dependent drift / diffusion.
4. Mixture models and Randomized Affine Diffusion (RAnD).
5. Calibration, simulation and pricing.
6. Conclusions.

Introduction

1. Background on xVAs:
   a. Economic value = risk-neutral value − xVA.
   b. Valuation Adjustments (xVAs), e.g., CVA, DVA, FVA, MVA, KVA.
   c. Computational challenges.
2. Focus on xVAs for IR derivatives.
3. Common industry xVA modeling setup in a Monte Carlo framework:
   a. Use one-factor short-rate model in Affine Diffusion class.
   b. Analytic tractability motivates use for xVA purposes.
   c. Example: Hull-White one-factor model (HW1F).

HW1F model

1. Impossible to fit to the whole market volatility surface (expiry $\times$ tenor $\times$ strike).
2. Time-dependent piece-wise constant volatility parameter used to calibrate the model to a strip of ATM co-terminal swaptions.
3. Forward rate under HW1F is shifted-lognormal: there is skew but it cannot be controlled.
4. The model does not generate volatility smile.
5. HW1F dynamics in the G1++ form:

$$r(t) = x(t) + b(t), \quad \mathrm{d}x(t) = -a_x x(t)\mathrm{d}t + \sigma_x(t)\mathrm{d}W_x(t).$$

Smile and skew: the market

1. Volatility smile on the short end.
2. Transforms into skew over time.

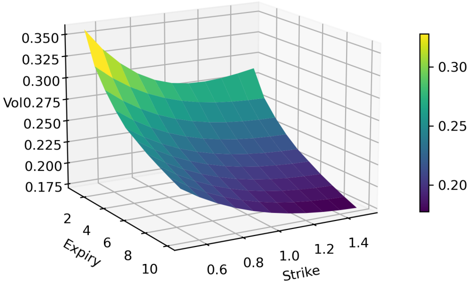
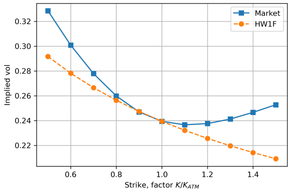
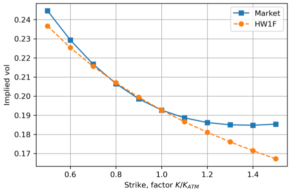


Figure: USD swaption volatility surface with 10Y tenor, market data from 28/09/2022. The volatilities are shifted Black volatilities. The strike is given as a factor times the ATM strike K_{ATM}, e.g., 1.2 means a strike of $1.2 \cdot K_{\text{ATM}}$.

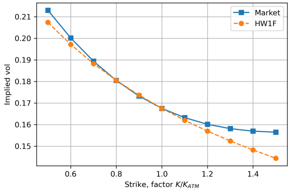
Smile and skew: the market vs HW1F



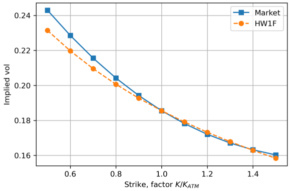
(a) 1Y expiry, 29Y tenor.



(b) 5Y expiry, 25Y tenor.



(c) 10Y expiry, 20Y tenor.



(d) 25Y expiry, 5Y tenor.

Figure: USD 30Y co-terminal swaption volatility strips (02/12/2022).

Smile and skew: xVA

1. Smile and skew typically absent from xVA calculations.
2. Challenge: find a model that captures smile and skew, but also allows for efficient calibration and pricing.
3. Smile and skew can be relevant for xVA:
   a. Obvious case: derivatives that take into account smile.
   b. Also for linear derivatives: legacy trades that are off-market and not primarily driven by ATM vols.
   c. Larger effect expected on PFE as this is a tail metric.
4. Examples in literature:
   a. Andreasen used a four-factor Cheyette model with local and stochastic volatility [**?**].
   b. Quadratic Gaussian models (quadratic form for the short rate) also allow smile control [**?**, Section 16.3.2].

Our contribution

1. Find SDE with state-dependent drift / diffusion that is consistent with the mixture of N different HW1F models, where one model parameter is varied.
2. This mixture model allows to capture market smile and skew
3. Profit from the analytic tractability of Affine Diffusion dynamics.
4. The model allows for fast and semi-analytic swaption calibration.
5. Monte Carlo pricing using regression methods.
6. Use the idea of the RAnD method to parameterize the mixture parameters.

SDE with state-dependent drift / diffusion

1. General dynamics for $r(t)$ for which we try to find the (potentially) state-dependent drift and diffusion:

$$\mathrm{d}r(t) = \mu_r(t, r(t))\mathrm{d}t + \eta_r(t, r(t))\mathrm{d}W_r(t). \quad (1)$$

2. We want to find $\mu_r(t, r(t))$ and $\eta_r(t, r(t))$ s.t. $\forall t$ the density is consistent with the weighted sum of N densities of analytically tractable models $r_n(t)$:

$$f_{r(t)}(y) = \sum_{n=1}^{N} \omega_n f_{r_n(t)}(y), \quad (2)$$

where

$$\mathrm{d}r_n(t) = \mu_{r_n}(t, r_n(t))\mathrm{d}t + \eta_{r_n}(t, r_n(t))\mathrm{d}W_{r_n}(t). \quad (3)$$

3. $\sum_{n=1}^{N} \omega_n = 1$ and $\omega_n > 0 \ \forall n$.
4. We derive $\mu_r(t, r(t))$ and $\eta_r(t, r(t))$ using the Fokker-Planck eq.

Fokker-Planck: applied to our case

We write down the FP equation for both $r(t)$ and $r_n(t)$. Using

$$f_{r(t)}(y) = \sum_{n=1}^{N} \omega_n f_{r_n(t)}(y), \tag{4}$$

and linearity of the derivative operator we obtain:

$$\mathrm{d}r(t) = \mu_r(t, r(t))\mathrm{d}t + \eta_r(t, r(t))\mathrm{d}W_r(t), \tag{5}$$

$$\mu_r(t, y) = \sum_{n=1}^{N} \mu_{r_n}(t, y)\Lambda_n(t, y), \tag{6}$$

$$\eta_r^2(t, y) = \sum_{n=1}^{N} \eta_{r_n}^2(t, y)\Lambda_n(t, y), \tag{7}$$

$$\Lambda_n(t, y) = \frac{\omega_n f_{r_n(t)}(y)}{\sum_{i=1}^{N} \omega_i f_{r_i(t)}(y)}. \tag{8}$$

The $r_n(t)$ dynamics

We choose to work with the HW1F model in the G1++ formulation, where each $r_n(t)$ has a different mean-reversion θ_n:

$$r_n(t) = x_n(t) + b_n(t), \tag{9}$$

$$\mathrm{d}x_n(t) = -\theta_n x_n(t)\mathrm{d}t + \sigma_x(t)\mathrm{d}W_x(t), \tag{10}$$

$$b_n(t) = f^{\mathrm{M}}(0,t) - x_n(0)\mathrm{e}^{-\theta_n t} + \int_0^t \sigma_x^2(u)B_n(u,t)\mathrm{e}^{-\theta_n(t-u)}\mathrm{d}u, \tag{11}$$

$$B_n(s,t) = \frac{1}{\theta_n}\left(1 - \mathrm{e}^{-\theta_n(t-s)}\right). \tag{12}$$

Here, $r_n(t) \sim \mathcal{N}\left(b_n(t) + \mathbb{E}_{t_s}[x_n(t)], \mathbb{V}\mathrm{ar}_{t_s}(x_n(t))\right)$ conditional on $\mathcal{F}_s$. Hence, we have that $f_{r_n(t)}(y)$ is a normal probability density function.

The $r_n(t)$ dynamics

Writing these dynamics in the desired form

$$\mathrm{d}r_n(t) = \mu_{r_n}(t, r_n(t))\mathrm{d}t + \eta_{r_n}(t, r_n(t))\mathrm{d}W_{r_n}(t) \tag{13}$$

yields

$$\begin{aligned}\mu_{r_n}(t, r_n(t)) &= \frac{\mathrm{d}\, f^{\mathrm{M}}(0,t)}{\mathrm{d}\, t} + \theta_n f^{\mathrm{M}}(0,t) - \theta_n r_n(t) \\ &\quad + \int_0^t \sigma_x^2(u)\mathrm{e}^{-2\theta_n(t-u)}\mathrm{d}u, \qquad (14)\\ \eta_{r_n}(t, r_n(t)) &= \sigma_x(t). \qquad (15)\end{aligned}$$

The $r(t)$ dynamics

Using these results, we have that

$$\mathrm{d}r(t) = \mu_r(t, r(t))\mathrm{d}t + \eta_r(t, r(t))\mathrm{d}W_r(t), \tag{16}$$

$$\mu_r(t, r(t)) = \sum_{n=1}^{N} \left[\frac{\mathrm{d}\, f^{\mathrm{M}}(0,t)}{\mathrm{d}\, t} + \theta_n f^{\mathrm{M}}(0,t) - \theta_n r(t) \right.$$
$$\left. + \int_0^t \sigma_x^2(u) \mathrm{e}^{-2\theta_n(t-u)} \mathrm{d}u \right] \Lambda_n(t, r(t)), \tag{17}$$

$$\eta_r(t, r(t)) = \sqrt{\sum_{n=1}^{N} \sigma_x^2(t) \Lambda_n(t, r(t))} = \sigma_x(t), \tag{18}$$

as $\sum_{n=1}^{N} \Lambda_n(t, y) = 1 \ \forall y$.
This means that the diffusion component $\eta_r(t, r(t))$ is unchanged, whereas the drift $\mu_r(t, r(t))$ is state-dependent.

Mixture models: general introduction

We derived the dynamics $X(t)$ that corresponds to the mixture of N different models $X_n(t)$:

1. The mixture is then driven by $f_{X(t)}(y)$:

$$f_{X(t)}(y) = \sum_{n=1}^{N} \omega_n f_{X_n(t)}(y). \tag{19}$$

2. A similar result holds for the valuation of a derivative $V_X(t)$:

$$V_X(t) = \sum_{n=1}^{N} \omega_n V_{X_n}(t). \tag{20}$$

3. (20) obtained for call option on equity when imposing (19) under the T-forward measure.
4. Equation (20) holds for non-path-dependent derivatives only.
5. For more complex derivatives, derive state-dependent (local-vol type) dynamics as before.

Mixture models: RAnD in general

Randomized Affine Diffusion (RAnD) method [**?**, **?**]:

1. Take an Affine Diffusion (AD) model.
2. Pick model parameter ϑ to randomize.
3. The random variable ϑ is defined on domain $D_\vartheta := [a, b]$ with PDF $f_\vartheta(x)$ and CDF $F_\vartheta(x)$, and realization θ, $\vartheta(\omega) = \theta$, such that the moments are finite.
4. For valuation, we use Gauss-quadrature weights $\{\omega_n, \theta_n\}_{n=1}^N$ where the nodes θ_n are based on $F_\vartheta(x)$, see [**?**, Appendix A.2]. Then, for valuation we can write:

$$V(t, r(t;\vartheta)) = \int_{[a,b]} V(t, r(t;\theta))\mathrm{d}F_\vartheta(\theta) \approx \sum_{n=1}^{N} \omega_n V(t, r(t;\theta_n)).$$

Mixture models: RAnD for mixture parametrization

1. Use the idea of the RAnD method to reduce dimensionality of the mixture parameters.
2. We do not suffer from the quadrature error.
3. We work with the HW1F dynamics.
4. We choose $\vartheta = a_x$, i.e., the mean-reversion parameter.
5. Impose $\mathcal{N}\left(\mu_\vartheta, \sigma_\vartheta^2\right)$ as randomizer (constant over time).
6. Key advantage: one additional degree of freedom w.r.t. HW1F.
7. $N = 5$ suitable when ϑ follows a normal (or uniform) distribution.

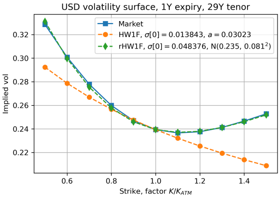
Calibration of the $r(t)$ dynamics

1. Calibration of the $r_n(t)$ dynamics in the usual way.
2. Mean-reversion mixture parameterized as $a_x \sim \mathcal{N}\left(\mu_\vartheta, \sigma_\vartheta^2\right)$. For each choice of μ_ϑ and σ_ϑ^2:
   a. Compute collocation points (Gauss-quad weights) $\{\omega_n, \theta_n\}_{n=1}^N$.
   b. Initialize N HW1F models with mean-reversion parameter $a_x = \theta_n$.
3. Use fast valuation

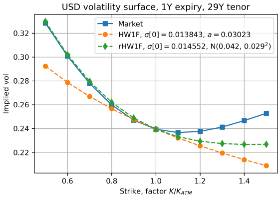
$$V(t, r(t;\vartheta)) = \sum_{n=1}^{N} \omega_n V(t, r(t;\theta_n)).$$

4. Calibrate the parametrization of the mean-reversion mixture $a_x \sim \mathcal{N}\left(\mu_\vartheta, \sigma_\vartheta^2\right)$ according to the desired strategy:
   a. Fit the initial coterminal smile.
   b. Fit all ATM points of the vol surface.
   c. Fit all coterminal smiles.
5. Bootstrap calibration of piece-wise constant model volatility to get a good ATM to the coterminal swaption strip.

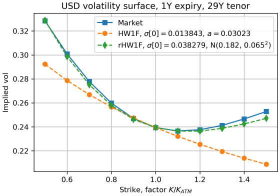
Calibration results



(a) Fit initial coterminal smile.



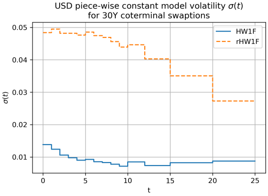
(b) Fit all ATM points.



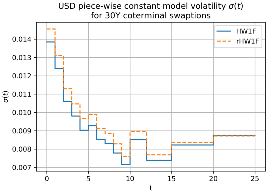
(c) Fit all coterminal smiles.

Figure: Initial coterminal smile. USD market data from 02/12/2022.

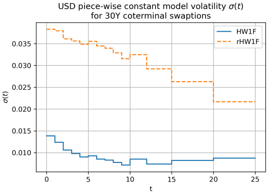
Calibration results



(a) Fit initial coterminal smile.



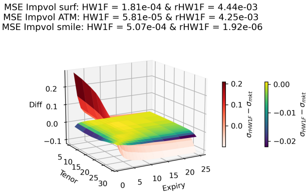
(b) Fit all ATM points.



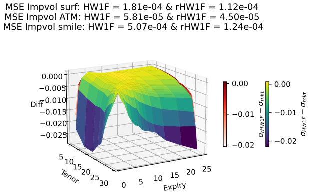
(c) Fit all coterminal smiles.

Figure: Calibrated model volatilities. USD market data from 02/12/2022.

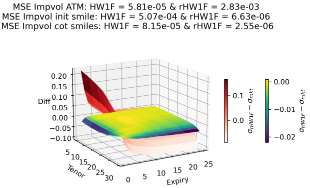
Calibration results



(a) Fit initial coterminal smile.



(b) Fit all ATM points.



(c) Fit all coterminal smiles.

Figure: Difference in ATM implied vols. USD market data from 02/12/2022.

Simulation of the $r(t)$ dynamics

Back to our dynamics:

$$\mathrm{d}r(t) = \mu_r(t, r(t))\mathrm{d}t + \eta_r(t, r(t))\mathrm{d}W_r(t), \tag{21}$$

$$\mu_r(t, r(t)) = \sum_{n=1}^{N} \left[\frac{\mathrm{d}\, f^{\mathrm{M}}(0,t)}{\mathrm{d}\, t} + \theta_n f^{\mathrm{M}}(0,t) - \theta_n r(t) + \int_0^t \sigma_x^2(u) \mathrm{e}^{-2\theta_n(t-u)} \mathrm{d}u \right] \Lambda_n(t, r(t)), \tag{22}$$

$$\eta_r(t, r(t)) = \sqrt{\sum_{n=1}^{N} \sigma_x^2(t) \Lambda_n(t, r(t))} = \sigma_x(t), \tag{23}$$

as $\sum_{n=1}^{N} \Lambda_n(t, y) = 1\ \forall y$.
This means that the diffusion component $\eta_r(t, r(t))$ is unchanged, whereas the drift $\mu_r(t, r(t))$ has become state-dependent.

Simulation of the $r(t)$ dynamics

1. We can always resort to an Euler-Maruyama discretization
$$r(t_{i+1}) = r(t_i) + \mu_r(t_i, r(t_i))\Delta t + \eta_r(t, r(t_i))\sqrt{\Delta t} Z, \qquad (24)$$
where $Z \sim \mathcal{N}(0,1)$.
2. Ideally we make large time steps. Hence, we integrate $\mathrm{d}r(t)$ to obtain an expression for $r(t)$ conditional on $r(s)$ for $s < t$, i.e.,
$$r(t) = r(s) + \int_s^t \mu_r(u, r(u))\mathrm{d}u + \int_s^t \eta_r(u, r(u))\mathrm{d}W_r(u). \qquad (25)$$
3. The integrated drift is difficult to compute:
$$\int_s^t \mu_r(u, r(u))\mathrm{d}u = f^{\mathrm{M}}(0,t) - f^{\mathrm{M}}(0,s)$$
$$+ \int_s^t \sum_{n=1}^N \left[\theta_n f^{\mathrm{M}}(0,u) - \theta_n r(u) + \mathbb{V}\mathrm{ar}_0\left(x_n(u)\right)\right] \Lambda_n(u, r(u))\mathrm{d}u.$$
4. Ideas: predictor-corrector method or Seven-League scheme.

Pricing under the $r(t)$ dynamics

1. Valuation as convex combination of underlying prices only for Europeans.
2. In general, we can use Monte Carlo with regression:
   a. Regression to avoid nested simulation.
   b. For example, we simulate r from t_0 to t and at this point we want to compute $P(t, T) = \mathbb{E}_t \left[\mathrm{e}^{-\int_t^T r(s)\mathrm{d}s} \right]$.
   c. For each T where we need $P(t, T)$ in pricing, it is regressed on $r(t)$.
   d. For example, an n-th order polynomial can be used as regression function, or something of exponential form.
3. These regression-based methods lend themselves naturally for xVA calculations, also known as American Monte Carlo.

Pricing a swaption under the $r(t)$ dynamics

1. Swaption with 10k notional, 5y expiry, on a 5y payer swap with annual payments.
2. Use 10^5 MC paths (antithetic variates turned on) and 100 simulation dates per year.
3. Polynomial regression of degree 4.

	Value	Imp.vol
HW1F: analytic	328.63814	0.22186
Convex comb: analytic	580.31577	0.40080
Convex comb: MC analytic ZCB	580.29341	0.40079
Convex comb: MC regressed ZCB	582.41497	0.40235
Mixture dynamics: MC regressed ZCB	581.20828	0.40146
Abs diff	1.20669	8.92e-04
Rel diff	2.08e-03	

Table: Results for all coterminal smiles calibration. Absolute and relative differences are between convex combination and mixture dynamics values using the MC with regressed ZCB. Mixture 95% conf.int.: $(578.96, 583.46)$.

Conclusions

1. Find SDE with state-dependent drift / diffusion that is consistent with the mixture of N different HW1F models, where one model parameter is varied.
2. This mixture model allows to capture market smile and skew
3. Profit from the analytic tractability of Affine Diffusion dynamics.
4. The model allows for fast and semi-analytic swaption calibration.
5. Monte Carlo pricing using regression methods.
6. Use the idea of the RAnD method to parameterize the mixture parameters.

Incorporating Smile in Valuation Adjustments Through the Mixture of Short-Rate Models

Utrecht University & Rabobank, the Netherlands

T. van der Zwaard

2023 SIAM Financial Mathematics and Engineering - June 7, 2023