Introduction to Interest Rate Models

Antoine Savine, Danske Bank, January-March 2018

Contents, part I

Contents, part II

1. The general Heath-Jarrow-Morton (HJM, 1992) framework
2. The Markov Property and the Cheyette (1992) model family
3. (Volatility, skew and kurtosis in Cheyette models)
4. The Multi-Factor Cheyette (MFC) model

A simple Interest Rate Model

The simplest dynamic IRM:
Ho-Lee (1986) under a Heath-Jarrow-Morton (1992) approach

IRM = model of the whole yield curve

- Black-Scholes (1973) and extensions (local volatility: Dupire, 1992 and stochastic volatility)
  - Given today's underlying asset price
  - Model its future arbitrage-free evolution (and value options and other contingent claims)
  - Applies when we model **one interest rate** (e.g. to price one swaption) but not when we model **all interest rates** (curve)
- In the context of multiple interest rates, what is the underlying asset?
  - Clean theoretical grounds for interest rate modelling introduced by Heath-Jarrow-Morton (HJM, 1992)
  - Underlying "asset" = whole curve = collection of all rates of all maturities at a given time
  - Today's **curve** is given and we model its future evolution
  - In order to price options and exotics, and estimate exposures
- Before HJM, we had short rate models where the short rate "drives" the complete curve
  - Main source: Hull-White (1990), (following Vasicek, 1977): assume short rate dynamics ➔ deduce curve dynamics
  - Also discrete short rate models (similar to binomial models on asset prices) by Ho-Lee (1986) and Black-Derman-Toy (1990)
  - HJM approach is superior and short term models have been abandoned by the industry
    - Theoretical foundations of short rate models are somewhat muddy
    - Short rate model parameters must be fitted to today's curve to prevent arbitrage
    - Unclear how to estimate and set volatility of short rate
    - Multi-factor extension possible (see Hull and White's two-factor extension) but somewhat unnatural
  - HJM models overcome all these limitations: directly model curve dynamics
  - Special HJM models (called Markov models) may be represented as short rate models (without their limitations) (Cheyette, 1992): assume curve dynamics ➔ deduce short rate dynamics
  - And (crucially) they naturally extend to multi-factor models (part II), contrarily to short rate models

Discount Factors (DF)

- For simplicity, we neglect spreads and credit and consider a single YC
- Yield Curve (YC) at time t = rates of all maturities T at time t = all discount factors of all maturities T at time t : DF(t,T)

- DF(t,T) = price at time t of 1 monetary unit paid at maturity T
- For a fixed maturity T, DF(t,T) is the price series of a tradable asset = zero-coupon bond of maturity T = delivers 1 at T

- At a time t, if we know all DF(t,T), we also know all the forward Libors and par swap rates:

$$L(t,T_1,T_2) = \frac{DF(t,T_1) - DF(t,T_2)}{(T_2 - T_1) DF(t,T_2)} \quad , FSR(t,T_1,T_2) = \frac{DF(t,T_1) - DF(t,T_2)}{\sum_{\text{fixed coupons}} \left(T^e{}_i - T^s{}_i\right) DF\left(t,T^e{}_i\right)}$$

(simplistic formulas, only correct in absence of spreads)

➔ The collection DF(t,T) represents all rate information at time t (again, in the absence of spreads)

Instantaneous Forward Rates (IFR)

- Another convenient representation of YC(t) is the collection of Instantaneous Forward Rates (IFR) of all maturities T at time t

- f(t,T) = seen at time t, par rate for a short term forward loan maturity T = (very) short term forward Libor maturity T

- Forward rates are deduced from discount factors and vice-versa

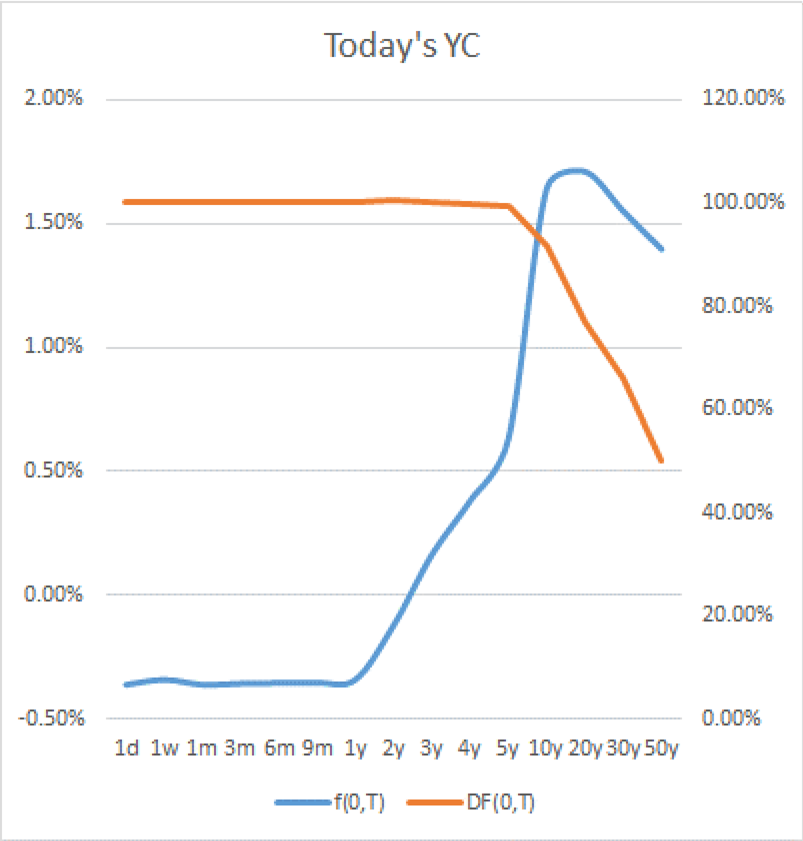
$$f\left(t,T\right) \equiv \lim_{\varepsilon \to 0} L\left(t,T,T+\varepsilon\right) = -\frac{\partial \log DF\left(t,T\right)}{\partial T} \Leftrightarrow DF\left(t,T\right) = \exp\left[-\int_t^T f\left(t,u\right) du\right]$$

- Note rates are **not** tradable assets, just a convenient “view” over bonds prices / discount factors

- One particular IFR: forward rate maturity t at time t = short rate at time t $r_t \equiv f\left(t,t\right)$

Today’s yield curve

- Todays Yield Curve YC(0)
  - Collection of all discount factors DF(0,T)
  - Or equivalently all forward rates f(0,T)
- In practice YC(0) is constructed out of linear instruments: swaps, coupon bonds, ...
- YC construction is a sophisticated process
- Out of scope for this presentation
- See for instance: History of Discounting, Savine, 2014 (available on slideShare)
- Here: we assume todays YC DF(0,T) and/or f(0,T) is known
- And model their future evolution



Yield Curve Model

- Todays YC(0) is given
- IRM specify how YC evolves from here

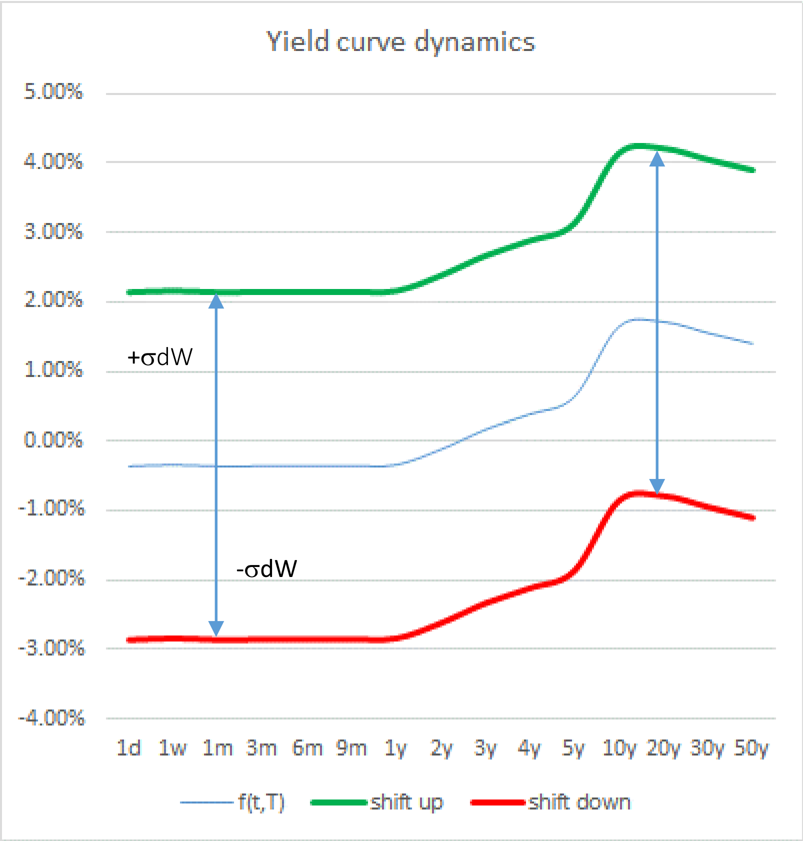
- Simplest model = parallel shifts = flat deformations

- Modelled with a Brownian Motion under the **historical** probability P: $df(t,T) = \sigma dW$

- (Constant) sigma = (annual) volatility of (all) rates

- Then all rates are normally distributed
  - with (same) (annual) variance sigma^2
  - And 100% correlation between rates of different maturities

- This model is far too simplistic for practical use
- But good to learn important IRM concepts in a simple context
- For instance, we will see that such dynamics is **impossible**



Convexity Arbitrage

- Dynamics of (tradable) bonds:
  Ito's lemma: 2nd order expansion of $DF(t,T) = \exp\left[-\int_t^T f(t,u)\,du\right]$

$$\frac{dDF(t,T)}{DF(t,T)} = d\log DF(t,T) + \frac{\left[d\log DF(t,T)\right]^2}{2}$$

$$DF(t,T) \equiv \exp\left[-\int_t^T f(t,u)\,du\right] \Rightarrow d\log DF(t,T) = -d\int_t^T f(t,u)\,du = r_t dt - \sigma(T-t)\,dW$$

$$\left[d\log DF(t,T)\right]^2 = \sigma^2 (T-t)^2\,dt$$

$$\boxed{\underbrace{\frac{dDF(t,T)}{DF(t,T)}}_{\text{bond return}} = \underbrace{r_t dt}_{\substack{\text{earns risk free rate}\\ \text{as maturity gets closer}}} \underbrace{-(T-t)\sigma dW}_{\text{volatility} \propto \text{duration}} + \underbrace{(T-t)^2 \frac{\sigma^2}{2} dt}_{\text{convexity} \propto \text{duration}^2}}$$

- Arbitrage = take advantage of convexity
  - Buy long term bond, for example 10y
  - Hedge with short term bond, for example 1y
  - Match durations, sell ~10 1y bonds for each 10y bond
  - End up with positive convexity and no risk

- Formally: buy T2 bonds, sell T1 bonds (T2 > T1)
  - Buy $DF(t,T_1)(T_1-t)$ bonds T2
  - Sell $DF(t,T_2)(T_2-t)$ bonds T1
  - Value at t: $\pi_t = DF(t,T_1)DF(t,T_2)(T_1-T_2) < 0$
  - Change in value: $d\pi_t = \underbrace{\left[r_t\pi_t - (T_1-t)(T_2-t)\frac{\sigma^2}{2}\underbrace{\pi_t}_{<0}\right]}_{>r_t\pi_t} dt + \underbrace{0dW}_{\text{no risk}}$

- We see that:
  - This portfolio has no risk
  - And earns more than risk free rate

- Means parallel shift dynamics is **impossible**
  - Fixed Income hedge funds and trading desks would execute and exhaust the arbitrage

- If random YC deformations are really parallel
  - There must be a **simultaneous steepening**
  - Causing long term bonds to decrease in value
  - And neutralize the arbitrage

Arbitrage Free Dynamics

- Remove arbitrage with average steepening
  - Modelled by **tenor dependent drift**:

$$df(t,T) = \sigma dW + \mu(T-t)dt$$

  - Can we compute the what drift (= speed of steepening) **exactly** neutralizes convexity arbitrage?

- Computation of the drift
  - Updated dynamics of bonds:

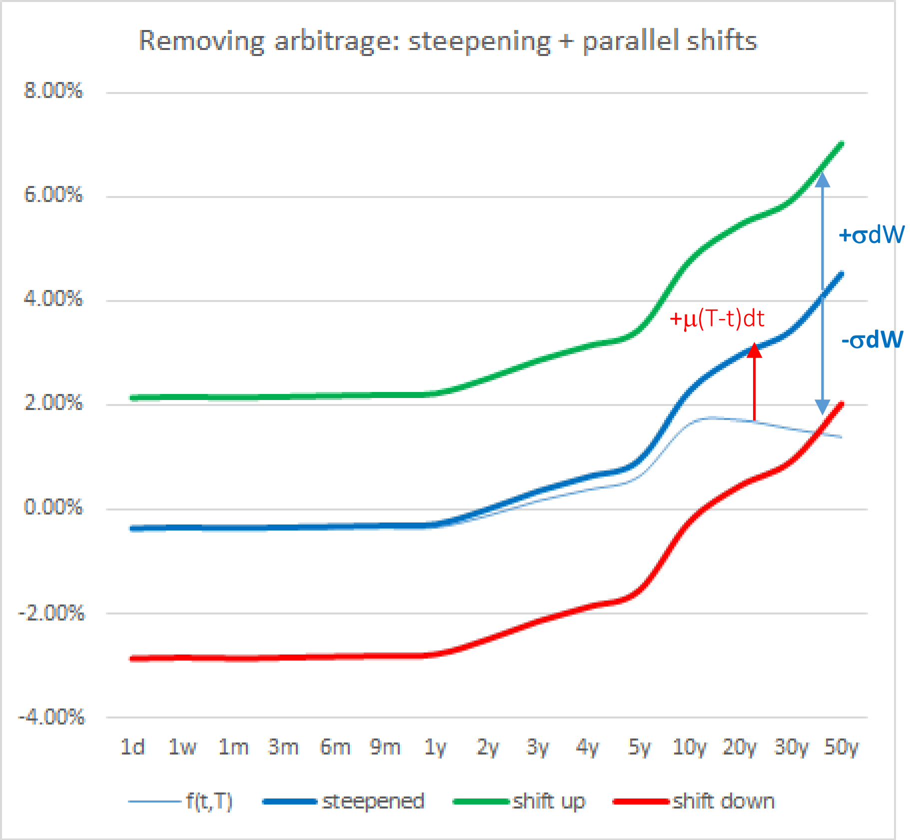
$$\frac{dDF(t,T)}{DF(t,T)} = r_t dt - (T-t)\sigma dW + (T-t)^2 \frac{\sigma^2}{2} dt - \left[\int_t^T \mu(T-u)du\right]dt$$

  - Updated dynamics of arbitrage portfolio:

$$\frac{d\pi_t}{dt} = r_t \pi_t + DF(t,T_1)DF(t,T_2)(T_1-t)(T_2-t)\left\{(T_2-T_1)\frac{\sigma^2}{2} - \left[\frac{\int_t^{T_2}\mu(T_2-u)du}{T_2-t} - \frac{\int_t^{T_1}\mu(T_1-u)du}{T_1-t}\right]\right\}$$

  - To prevent arbitrage:

$$\frac{d\pi_t}{dt} = r_t \pi_t \Leftrightarrow \frac{\frac{\int_t^{T_2}\mu(T_2-u)du}{T_2-t} - \frac{\int_t^{T_1}\mu(T_1-u)du}{T_1-t}}{T_2-T_1} = \frac{\sigma^2}{2} \quad (\forall T_2 > T_1) \text{ hence } \mu(T-t) = \sigma^2(T-t) + c(t)$$



Arbitrage-Free Dynamics (2)

- Under the **historical** probability, in the simple flat deformation model, changes in forward rates **must** satisfy

$$df(t,T)=\sigma dW+\left[\sigma^2(T-t)+c(t)\right]dt \text{ define } \eta_t\equiv-\frac{c(t)}{\sigma} \text{ and get } \boxed{df(t,T)=\Bigg[\underbrace{\sigma^2(T-t)}_{\text{convexity arbitrage adjustment}}-\underbrace{\eta_t\sigma}_{\text{risk premium}}\Bigg]dt+\underbrace{\sigma dW}_{\text{random parallel shifts}}}$$

  - The quantity η is called **risk premium**
  - It could depend on time t and even have stochastic dynamics
  - **But it must be the same for all rates of all maturities T**

- The bond dynamics is: $\boxed{\frac{dDF(t,T)}{DF(t,T)}=\left[r_t+\eta_t\sigma(T-t)\right]dt-\sigma(T-t)dW}$
  - Hence the risk premium is excess return per unit of (bond) volatility
  - And we reiterate that risk premium must be **the same for all bonds**

- To find the short rate dynamics:
  1. First integrate forward rate
  2. Find integrated form for r
  3. Differentiate

1. $$f(t,T)=f(0,T)+\sigma^2 t\left(T-\frac{t}{2}\right)-\sigma\int_0^t \eta_s ds+\sigma W_t$$

2. $$r_t=f(t,t)=f(0,t)+\frac{\sigma^2t^2}{2}-\sigma\int_0^t \eta_s ds+\sigma W_t$$

3. $$\boxed{dr_t=\Bigg[\underbrace{\frac{\partial f(0,t)}{\partial t}}_{\text{follow forwards}}+\underbrace{\sigma^2 t}_{\text{convexity adjustment}}-\underbrace{\sigma\eta_t}_{\text{risk premium}}\Bigg]dt+\sigma dW}$$

Markov Property

- From the integrated equation on forward rates $f(t,T) = \underbrace{f(0,T) + \sigma^2 t\left(T - \frac{t}{2}\right)}_{\substack{\text{deterministic} \\ \text{depends on T}}} \underbrace{-\sigma\int_0^t \eta_t dt + \sigma W_t}_{\substack{\textit{random} \\ \text{does not depend on T} \\ \text{common to all forwards}}}$

- It follows that $f(t,T_2) - f(t,T_1) = f(0,T_2) - f(0,T_1) + \sigma^2 t(T_2 - T_1)$ and $f(t,T_2) = f(t,T_1) + f(0,T_2) - f(0,T_1) + \sigma^2 t(T_2 - T_1)$

- **All forward rates are deterministic functions of one another – <u>and that function does not depend on risk premium</u>**

  - This means that the dynamics of the entire YC can be reduced to the dynamics of **one arbitrary forward rate** of some maturity T*

$$df(t,T^*) = \left[\sigma^2(T^* - t) - \eta_t \sigma\right] dt + \sigma dW$$

  - And all other rates at time t are found as a function of f(t,T*) with the **reconstruction formula** $f(t,T) = f(t,T^*) + f(0,T) - f(0,T^*) + \sigma^2 t(T - T^*)$

  - For example, we can reduce the whole YC dynamics to the dynamics of the short rate: $dr_t = \left[\frac{\partial f(0,t)}{\partial t} + \sigma^2 t - \sigma\eta_t\right] dt + \sigma dW$

  - And reconstruct the whole future YC as a function of the short rate alone: $f(t,T) = r_t + \underbrace{f(0,T) - f(0,t)}_{\text{today's slope}} + \underbrace{\sigma^2 t(T - t)}_{\text{convexity adjustment}}$

- Note: it follows that bond prices are also a deterministic function of one another or (or short or some forward rate)
(We compute that function later)

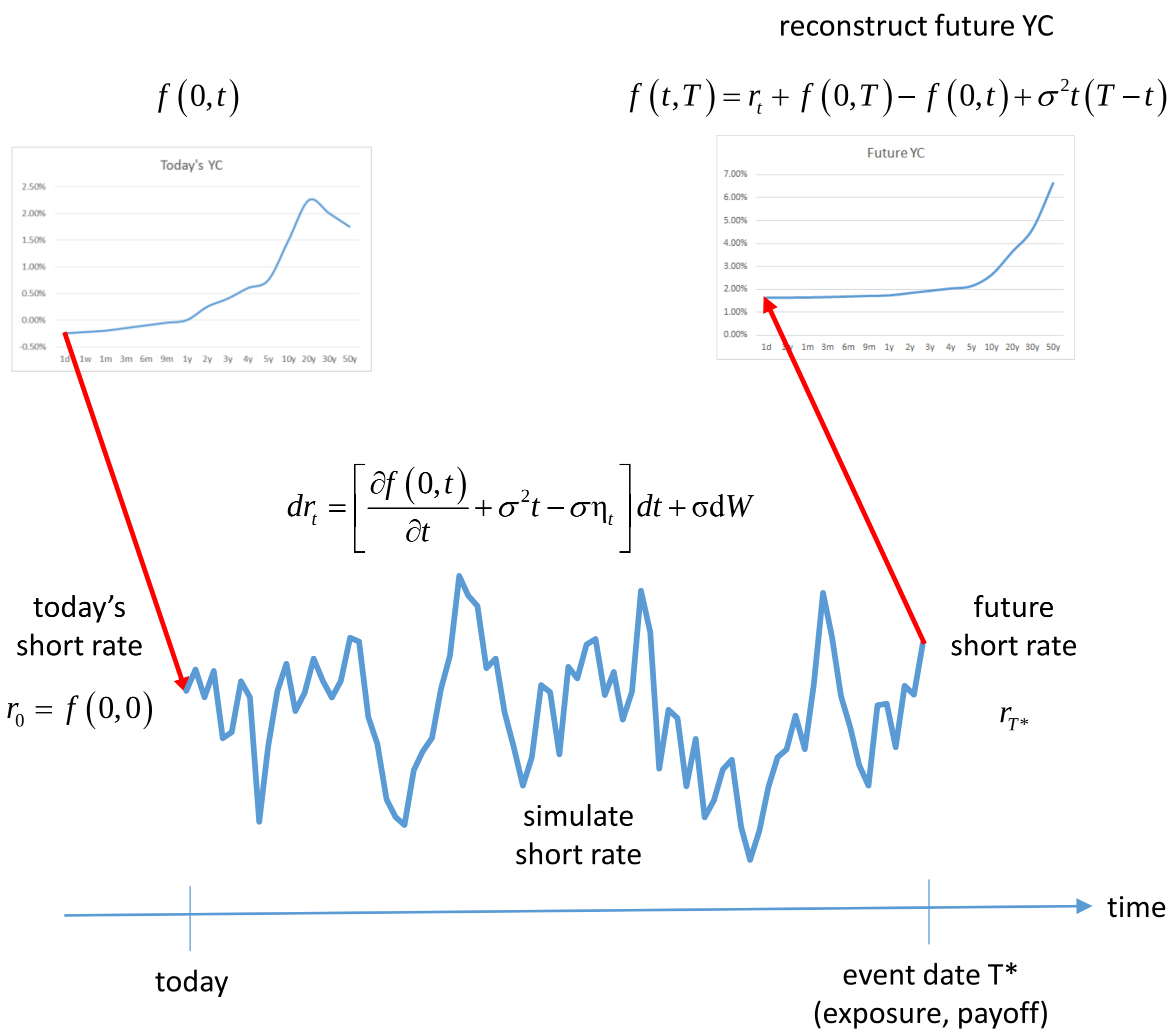
From curve to short rate model and back

- Change of variable: $X_t \equiv r_t - f(0,t)$
  - X: “random factor”, distance of realized short rate from forward
  - Then we have the factor dynamics
    $$dX_t = \left[\sigma^2 t - \sigma\eta_t\right]dt + \sigma dW$$
  - And reconstruction
    $$f(t,T) = \underbrace{f(0,T)}_{\text{today's forward}} + \underbrace{X_t}_{\text{common factor to all rates}} + \underbrace{\sigma^2 t(T-t)}_{\text{convexity adjustment}}$$

- Crucial property for an efficient implementation
  - No need simulate the whole YC
  - Only simulate factor X
  - And reconstruct the whole future YC as a known function of X

- Equivalence between short rate model and YC model
  - Short rate at time t encapsulates the whole curve at time t

- <u>Markov property **not** satisfied for general YC models</u>
  - Satisfied in the simplistic case with only parallel shifts
  - Later we identify all YC models that satisfy the Markov property



Simple parallel IRM: Take Away

- With parallel shifts, the dynamics or rates **under the historical probability** must be:

$$df(t,T) = \underbrace{\sigma^2 (T-t)dt}_{\text{deterministic steepening}} - \underbrace{\eta_t \sigma dt}_{\text{Rrisk premium}} + \underbrace{\sigma dW}_{\text{random parallel shift}}$$

σ: volatility of (all) rates, η_t: risk premium for (all) rates

- The risk premium must be the same for all rates, <u>independently of their maturity T</u>
- We have one risk premium by **factor** (one in our model), but same for all assets (bonds)
- It is the unicity of the risk premium **under the historical probability** that makes the dynamics arbitrage-free

- The induced dynamics on bonds is:

$$\underbrace{\frac{\mathrm{dDF}(\mathrm{t,T})}{\mathrm{DF}(\mathrm{t,T})}}_{\text{bond return}} = \underbrace{r_t dt}_{\text{earns short rate as maturity approaches}} + \underbrace{\eta_t \sigma (T-t) dt}_{\text{risk premium}} - \underbrace{\sigma (T-t)}_{\text{volatility} \propto \text{duration}} dW$$

- That model satisfies the Markov property: all rates are deterministic functions of one another

  - The model is identical to the short rate model $dr_t = \left[\frac{\partial f(0,t)}{\partial t} + \sigma^2 t - \sigma \eta_t\right] dt + \sigma \mathrm{d}W$ $\quad X_t \equiv r_t - f(0,t) \Rightarrow dX_t = \left(\sigma^2 t - \sigma \eta_t\right) dt + \sigma \mathrm{d}W$
  - With the reconstruction formula $f(t,T) = r_t + f(0,T) - f(0,t) + \sigma^2 t (T-t)$

$$f(t,T) = \underbrace{f(0,T)}_{\textit{forward}} + \underbrace{X_t}_{\textit{factor}} + \underbrace{\sigma^2 t (T-t)}_{\textit{convexity}}$$

➔ Today's YC is given, the model is parameterized by volatility and risk premium

Risk Premium
and Risk Neutralization

Pricing derivatives and conducting regulatory calculations under the parallel IRM

Derivatives Pricing

- Consider a European option delivering at time Tex a payoff dependent on YC(Tex)
  - For example, a swaption or a caplet or a coupon bond option
  - (Similar arguments apply to exotics with same results, although more complicated)
  - Denote v_t the value of the option at time t

- The value at time t is a function of the curve at time t
  - $v_t = g\left(t, YC_t\right)$
  - Can be proved, although formal mathematical demonstration rather abstract
  - Intuitively: the state of the world is represented by the YC in our simple model, everything at time t is a function of YC(t)
  - Note: Black and Scholes originally **postulated** that a call price at time t is function of the underlying asset price at time t formally demonstrated in the 1980s (Harrisson-Pliska and Harrisson-Kreps) for Europeans and 2009 (Dupire's functional Ito calculus) for exotics!

- From the Markov property
  - The whole YC(t) is a deterministic function of one bond price, we choose a bond of maturity T* > Tex: $YC_t = h\left[DF\left(t, T^*\right)\right]$
  - Hence $v_t = v\left[t, DF\left(t, T^*\right)\right]$

- Hence the dynamics of the option price is (by Ito applied to v):

$$dv_t = \frac{\partial v}{\partial t}dt + \frac{\partial v}{\partial DF}dDF\left(t,T^*\right) + \frac{\partial^2 v}{2\partial DF^2}\left(dDF\left(t,T^*\right)\right)^2 = \left[\frac{\partial v}{\partial t} + \frac{\partial v}{\partial DF}DF\left(t,T^*\right)\mu_{DF}\left(t,T^*\right) + \frac{\partial^2 v}{2\partial DF^2}DF\left(t,T^*\right)^2\sigma_{DF}\left(t,T^*\right)^2\right]dt + \frac{\partial v}{\partial DF}DF\left(t,T^*\right)\sigma_{DF}\left(t,T^*\right)dW$$

where μ_{DF} and σ_{DF} are the drifts and volatilities of bonds: $\mu_{DF}\left(t,T\right) = r_t + \eta_t \sigma\left(T-t\right)$, $\sigma_{DF}\left(t,T\right) = \sigma\left(T-t\right)$

Derivatives Pricing (2)

- Assuming the value of the option is strictly positive, we can write the dynamics of the option price in proportional terms:

$$\frac{dv_t}{v_t} = \left\{ \underbrace{\left[\frac{\partial v}{v_t \partial t} + \frac{\partial v}{\partial DF} \frac{DF(t,T^*)}{v_t} \mu_{DF}(t,T^*) + \frac{\partial^2 v}{2\partial DF^2} \frac{DF(t,T^*)^2}{v_t} \sigma_{DF}(t,T^*)^2 \right]}_{\text{(historical) drift of the option price} = \mu^v_t} \right\} dt + \underbrace{\frac{\partial v}{\partial DF} \frac{DF(t,T^*)}{v_t} \sigma_{DF}(t,T^*)}_{\text{volatility of the option price} = \sigma^v_t} dW$$

- We can hedge the risk of v at time t by selling $\frac{v_t \sigma^v_t}{DF(t,T^*)\sigma_{DF}(t,T^*)}$ bonds:
  - The resulting portfolio has 0 volatility between t and t+dt
  - Its value is: $\pi_t = v_t - \frac{v_t \sigma^v_t}{DF(t,T^*)\sigma_{DF}(t,T^*)} DF(t,T^*) = \left[1 - \frac{\sigma^v_t}{\sigma_{DF}(t,T^*)}\right] v_t$
  - And its drift is: $E_t\left[\frac{d\pi_t}{\pi_t dt}\right] = \frac{d\pi_t}{\pi_t dt} = \left[\mu^v_t - \frac{\sigma^v_t}{\sigma_{DF}(t,T^*)} \mu_{DF}(t,T^*)\right] \Big/ \left[1 - \frac{\sigma^v_t}{\sigma_{DF}(t,T^*)}\right]$
  - Arbitrage dictates that: $\frac{d\pi_t}{\pi_t dt} = r_t \Leftrightarrow \mu^v_t = \left[1 - \frac{\sigma^v_t}{\sigma_{DF}(t,T^*)}\right] r_t + \frac{\sigma^v_t}{\sigma_{DF}(t,T^*)} \mu_{DF}(t,T^*)$
  - And since $\mu_{DF}(t,T^*) = r_t + \eta_t \sigma_{DF}(t,T^*)$ it follow that $\mu^v_t = r_t + \eta_t \sigma^v_t$

Pricing under the Historical Probability

- We derived that: $\mu^{v}{}_{t} = r_t + \eta_t \sigma^{v}{}_{t}$
  - Means that the option is subject to **the same risk premium as the bonds**
  - The option earns an average (under historical probability) excess over risk-free rate of risk premium times its own volatility

- The dynamics of the option price simplifies into: $\frac{dv_t}{v_t} = \left(r_t + \eta_t \sigma^{v}{}_{t}\right) dt + \sigma^{v}{}_{t} dW$
  - (Although its volatility is complicated and depends on the yet unresolved pricing function and its derivatives)
  - A classical result in stochastic calculus states that: $\frac{dX_t}{X_t} = \mu_t dt + \sigma_t dW \Rightarrow X_{T_1} = E_{T_1}\left[\exp\left(-\int_{T_1}^{T_2} \mu_t dt\right) X_{T_2}\right]$
  - Hence: $\boxed{v_0 = E\left\{\exp\left[-\int_0^t \left(r_s + \eta_s \sigma^{v}{}_{s}\right) ds\right] v_t\right\} = E\left\{\exp\left[-\int_0^{T_{ex}} \left(r_t + \eta_t \sigma^{v}{}_{t}\right) dt\right] v_{T_{ex}}\right\} \text{ where } v_{T_{ex}} \text{ is the payoff}}$

- The price is the expected (under **historical** probability) payoff, discounted **at a risk-adjusted rate**
  - That rate is the risk free rate + risk premium times option’s own volatility
  - Hence, we re-demonstrated **through arbitrage arguments** a fundamental result from microeconomics:
    asset prices are expectations of their future values, discounted with risk adjusted rates
  - That property holds for all assets: primary (bonds) and derivatives (including exotics, although we only demonstrated Europeans)

Pricing under the Risk-Neutral Probability

- That formula: $v_0 = E\left\{\exp\left[\left(-\int_0^{T*}\left(r_t + \eta_t \sigma^v{}_t\right)dt\right)\right]v_{T*}\right\}$
    - Is important theoretically because it shows that derivatives can be priced under the historical probability
    - But it is unpractical:
        - Requires an estimation of the risk premium
        - And the (recursive) simulation of the option's volatility
- Remember we used the Markov property to show that: $v_t = v\left[t, DF\left(t,T*\right)\right]$
    - Price the option = find that function v
    - We derived that by Ito: $\mu^v{}_t = \frac{\partial v}{v_t \partial t} + \frac{\partial v}{\partial DF}\frac{DF\left(t,T*\right)}{v_t}\mu_{DF}\left(t,T*\right) + \frac{\partial^2 v}{2\partial DF^2}\frac{DF\left(t,T*\right)^2}{v_t}\sigma_{DF}\left(t,T*\right)^2$, $\sigma^v{}_t = \frac{\partial v}{\partial DF}\frac{DF\left(t,T*\right)}{v_t}\sigma_{DF}\left(t,T*\right)$
    - And we know that: $\mu_{DF}\left(t,T*\right) = r_t + \eta_t \sigma\left(T* - t\right)$, $\sigma_{DF}\left(t,T*\right) = \sigma\left(T* - t\right)$
    - And we have the arbitrage condition: $\mu^v{}_t = r_t + \eta_t \sigma^v{}_t$
- These combined equations resolve into: $\frac{\partial v}{\partial t} + \frac{\partial v}{\partial DF}DF\left(t,T*\right)r_t + \frac{\partial^2 v}{2\partial DF^2}DF\left(t,T*\right)^2\sigma^2\left(T* - t\right)^2 = r_t v$

Pricing under the Risk-Neutral Probability (2)

- We found that the price of the option $v_t = v\left[t, DF\left(t,T*\right)\right]$ is the solution of the Partial Differential Equation (PDE):

$$\frac{\partial v}{\partial t} + \frac{\partial v}{\partial DF} DF\left(t,T*\right) r_t + \frac{\partial^2 v}{2\partial DF^2} DF\left(t,T*\right)^2 \sigma^2 \left(T*-t\right)^2 = r_t v \text{ with boundary condition } v\left[T_{ex}, DF\left(T_{ex},T*\right)\right] = payoff$$

- This PDE is **independent of the risk premium** (simplified away during calculations)
  Why? We build a risk free portfolio by matching volatilities, which also removes risk premium in drift, proportional to volatility

- Hence the pricing function v is the same whatever the risk premium

- Since this is the case, we may as well use 0 risk premium when pricing

  - This means the dynamics of the YC is: $df\left(t,T\right) = \sigma^2\left(T-t\right)dt + \sigma dW$

  - And the pricing equation is: $v_0 = E\left[\exp\left(-\int_0^{T*} r_t dt\right) v_{T*}\right]$

- We can zero the risk premium for pricing because the pricing function is independent from it
  (risk premium changes expectations and discounting, and the 2 effects always compensate exactly)

$$E^{df\left(t,T\right)=\left[\sigma^2\left(T-t\right)-\eta_t\sigma\right]dt+\sigma dW}\left\{\exp\left[\left(-\int_0^{T*}\left(r_t+\eta_t\sigma^v{}_t\right)dt\right)\right]v_{T*}\right\} = E^{df\left(t,T\right)=dt+\sigma dW}\left\{\exp\left[\left(-\int_0^{T*} r_t dt\right)\right]v_{T*}\right\}$$

- That simplifies the pricing problem, in particular price only depends on YC(0) and volatility

Example: bond pricing

- Example: t value of the T bond = “option” that pays 1 at time T
  - We can price it under the historical probability: $DF(t,T)=E_t\left\{\exp\left[-\int_t^T\left[r_s+\eta_s\sigma_{DF}(s,T)\right]ds\right]\right\}=E_t\left\{\exp\left[-\int_t^T\left[r_s+\eta_s\sigma(T-s)\right]ds\right]\right\}$
  - With: $dr_t=\left[\frac{\partial f(0,t)}{\partial t}+\sigma^2 t-\sigma\eta_t\right]dt+\sigma dW\Rightarrow r_s=r_t+\left[f(0,s)-f(0,t)\right]+\sigma^2\frac{s^2-t^2}{2}+\sigma(W_s-W_t)-\sigma\int_t^s\eta_u du$
  - It follows that: $DF(t,T)=E_t\left\{\exp\left\{-\int_t^T\left[\underbrace{r_t+\left[f(0,s)-f(0,t)\right]+\sigma^2\frac{s^2-t^2}{2}+\sigma(W_s-W_t)-\sigma\int_t^s\eta_u du}_{r_s}+\eta_s\sigma(T-s)\right]ds\right\}\right\}$

$$=\exp\left\{-(T-t)\underbrace{\left[r_t-f(0,t)\right]}_{\text{factor } X_t}\right\}\underbrace{\exp\left[-\int_t^T f(0,s)ds\right]}_{=\frac{DF(0,T)}{DF(0,t)}}\underbrace{\exp\left[-\frac{\sigma^2 t(T-t)^2}{2}\right]}_{\text{convexity adjustment}}E_t\left\{\exp\left[\sigma\left(\underbrace{\int_t^T ds\int_t^s\eta_u du}_{\text{from the dynamics of r}}-\underbrace{\int_t^T(T-s)\eta_s}_{\text{from discounting}}\right)\right]\right\}$$

- Integration by parts shows that: $\int_t^T ds\int_t^s\eta_u du-\int_t^T(T-s)\eta_s=0$ discounting r.p. exactly compensates short rate dynamics r.p.

- So the value of the Bond is independent from r.p. and the bond price is the same under the risk-neutral measure:

$$\eta_t=0\Rightarrow dr_t=\left[\frac{\partial f(0,t)}{\partial t}+\sigma^2 t\right]dt+\sigma dW \text{ and } DF(t,T)=E_t\left\{\exp\left[-\int_t^T r_s ds\right]\right\}=\frac{DF(0,T)}{DF(0,t)}\exp\left[-(T-t)X_t-\frac{\sigma^2 t(T-t)^2}{2}\right]$$

Counter example: expected future bond price

- Another example: expectation of the T1 value of the T2 bond (T2>T1)
  - We know that: $f(T_1, T) = f(0,T) + \sigma^2 T_1 \left(T - \frac{T_1}{2} \right) - \sigma \int_0^{T_1} \eta_s ds + \sigma W_{T_1}$
  - Hence: $E\left[DF(T_1, T_2)\right] = E\left\{ \exp\left[-\int_{T_1}^{T_2} f(T_1, u) du \right] \right\}$

$$= \frac{DF(0,T_2)}{DF(0,T_1)} \exp\left[-\frac{\sigma^2 T_1 T_2 (T_2 - T_1)}{2} \right] E\left\{ \exp\left[\sigma (T_2 - T_1) \left(\int_0^{T_1} \eta_s ds - W_{T_1} \right) \right] \right\}$$

$$\underset{\text{if } \eta \perp W}{=} \underbrace{\frac{DF(0,T_2)}{DF(0,T_1)}}_{\text{forward}} \underbrace{\exp\left[\frac{\sigma^2 T_1^2 (T_2 - T_1)}{2} \right]}_{\text{convexity adjustment}} \underbrace{E\left\{ \exp\left[\sigma (T_2 - T_1) \int_0^{T_1} \eta_s ds \right] \right\}}_{\text{risk premium}}$$

- This illustrates that **expectations** of future values depend on risk premium

- However, **prices** (= future values <u>discounted with risk adjusted rates</u>) are independent of risk premium

- For instance, the **discounted** (with risk adjusted rate) expected value at T1 of the T2 bond is (obviously):

$$E\left\{ \exp\left[-\int_0^{T_1} \left(r_s + \eta_s \sigma (T_2 - t) \right) ds \right] DF(T_1, T_2) \right\} = DF(0, T_2)$$

Risk-Neutralization: Take Away

- The price of an option is its expected payoff **under the historical probability**, discounted with a **risk-adjusted** rate
- That risk adjusted rate = risk-free rate + risk premium times the volatility of the option price

- We demonstrated that the price is independent of risk premium, hence:
  - It is also the expected payoff, **discounted by the risk-free rate**
  - Under a dynamics where the risk premium is zero (for all assets)

- This is easier to calculate
  - No need to estimate or postulate risk premium
  - No need to discount with a rate adjusted by the option's volatility
  - Swaptions, caps and coupon bond options admit Black-Scholes like closed form formulas under this model, see .e.g. Jamshidian, 1989

- This is referred to as **pricing under the risk-neutral probability**
- Means we price "as if" all asset values, including options, are expected future values discounted at the risk-free rate
- We can do this because prices are independent of risk premium
- But this is nothing more than a calculation facility, it does not change the result

- Only prices = risk discounted future values are independent of risk premium, **undiscounted expectations depend on risk premium**

Risk premium and regulatory calculations

- **For pricing**, we can safely “move the problem to a risk neutral world”, but what about regulatory calculations like CCR?

- **For regulations**, in principle, we compute expected exposures
- These are (undiscounted) expectations, under the historical probability, not prices

- We know that expectations depend on risk premium
- That would force us to estimate the risk premium, a difficult and somewhat arbitrary exercise

- Interestingly, the regulator seems to be increasingly encouraging computation of **risk-neutral** expectations:
    - PRIIPS regulation explicitly requires risk neutral adjustment of historical data used for simulations (annex 2, paragraph 22, alinea c)
    - CCR allows risk-neutral simulations (act 292-2)
- Hence risk-neutralization seems to also apply to regulatory calculations, although for different reasons
- We believe the regulator wants to mitigate arbitrary setting/estimation of risk premium / asset price drift

- With risk premium set to 0, volatility is the only degree of freedom in the model

Yield Curve Models and Short Rate Models

Another look at the Markov property

- We have demonstrated the fundamental pricing formula under historical probability P: $v_t = E_t\left\{\exp\left[-\int_t^T \left(r_s + \eta_s \sigma^v{}_s\right) ds\right] v_T\right\}$

- We demonstrated that we obtain the same price under the risk-neutral probability Q, discounting at the risk-free rate: $v_t = E^Q{}_t\left[\exp\left(-\int_t^T r_s ds\right) v_T\right]$

- In particular, for a bond of maturity T: $DF(t,T) = E^Q{}_t\left[\exp\left(-\int_t^T r_s ds\right)\right]$

- In the theory of Stochastic Processes, a process X is Markov when:
  - The conditional expectation at t, of any functional g of the future path of X, only depends on Xt $E_t\left[g\left(X_s, t \le s\right)\right] = h\left(t, X_t\right)$
  - “The future only depends on the past through the present”

- An important result demonstrates that any self-contained diffusion of the type $dX_t = \mu\left(t, X_t\right)dt + \sigma\left(t, X_t\right)dW$ is Markovian

- In our simple (parallel) model, the short rate is a self contained diffusion under Q (but not under P due to risk premium!) $dr_t = \left[\frac{\partial f\left(0,t\right)}{\partial t} + \sigma^2 t\right]dt + \sigma dW^Q$
  - This is not true of all interest rate models!
  - In our model, and all models where the short rate is a Q diffusion, hence Markovian, it follows from $DF(t,T) = E^Q{}_t\left[\exp\left(-\int_t^T r_s ds\right)\right]$ that
    - All bond prices at time t are deterministic functions of the short rate at time t
    - The model can be reduced to a short rate model

From yield curve models to short rate models

- We followed the methodology established by Heath-Jarrow-Morton (HJM, 1992):
  - We started with today's yield curve
  - And modelled its future evolution
  - By specifying its **deformations** (for now, parallel)

- We proved that, in this case, the dynamics of all forward rates **must** satisfy:

$$df(t,T) = \Bigg[\underbrace{\sigma^2 (T-t)}_{\text{convexity arbitrage adjustment}} - \underbrace{\eta_t \sigma}_{\text{risk premium}} \Bigg] dt + \underbrace{\sigma dW}_{\text{random parallel shifts}}$$

- In particular, we have the dynamics of the short rate:

$$dr_t = \Bigg[\underbrace{\frac{\partial f(0,t)}{\partial t}}_{\text{follow forwards}} + \underbrace{\sigma^2 t}_{\text{convexity adjustment}} - \underbrace{\sigma \eta_t}_{\text{risk premium}} \Bigg] dt + \sigma dW$$

- We then used the Markov property to show that all rates in the future are a known function of the short rate:

$$f(t,T) = r_t + \underbrace{f(0,T) - f(0,t)}_{\text{today's slope}} + \underbrace{\sigma^2 t (T-t)}_{\text{convexity adjustment}}$$

- We have effectively rewritten the yield curve model as a short rate model

Short rate model parameters

- The short rate model is only a convenient **representation** of the yield curve model
- That has consequences for the parameterization of the model
- In our simple model, the only parameter is sigma
    - Sigma is the volatility of the short rate
    - It is also the amplitude of yield curve deformations
- If we estimate sigma historically
    - Prices and exposures typically depend on the volatility of long (swap) rates
    - It would be a mistake to set sigma to the (historical) volatility of the short rate
    - It should be set to the average amplitude of curve deformations
- For example in 2017
    - Volatility of the short rate was very low (less than 5 bppa for the 3m rate)
    - But volatility of longer rates was substantially higher (around 50 bppa for the 10y rate)
    - Setting sigma to the short rate volatility would severely underestimate swap rate volatility
- More generally (part II)
    - IRM parameters = shape and amplitude of yield curve deformations
    - They must be set accordingly
    - And not from the short rate model representation

Historical and Implied Volatility

The Fundamental Theorem of Derivatives Pricing (Rolf Poulsen, 2015)

Historical volatility estimation: choice of estimation window

- We highlight problems with historical estimation

- One is choice of estimation window
  - Example: S&P volatility as of Dec 2017
  - 1Y = 6.75%, 2Y = 10.50%, 3Y = 12.50%, 5Y = 12%, 10Y = 20%
  - Which is a reasonable predictor for 2018?

- Advanced statistical methods exist to try answer that question
  - Exponential weighting
  - ARCH/GARCH
  - Basically estimate volatility as a mean-reverting process towards long term equilibrium

- Alternatively, use implied volatility
  - Forward looking
  - Market’s best estimate of future realized (see next)
    - Implicitly, consensus from professional traders
    - Willing to wage monetary stakes on that forecast

Historical volatility estimation: past and future events

- Sensitivity to past events
  - Example: EUR/CHF unpegging, 15^{th} January 2015
  - Produced extreme volatility
  - A one-off event, not repeatable in the foreseeable future
  - EUR/CHF volatility: 18% over 2015, 5% over 2016
  - Clearly, estimation over 2015 a poor predictor for 2016
- Ignorance of future events
  - Example: ECB easing from January 2015
  - Increased EUR/USD volatility
  - Known as almost certain late 2014
  - Yet not reflected in historical estimate
  - EUR/USD volatility: 7% June-Dec 2014, 13% Jan-June 2015
- Voluntarily extreme examples that demonstrate defects of naïve estimation
- Methods exist to adjust historical estimates for events
  - Clear impact of past events before estimation
  - Add predicted impact of future events after estimation
- Implied volatility = forward looking estimate, naturally adjusted for past and future events

Historical volatility estimation: black swans

- Black swan events
  - Example: the global crisis of 2008
  - By definition, very rare events (3 major crises in the past 100 years)
  - With massive impact on volatility

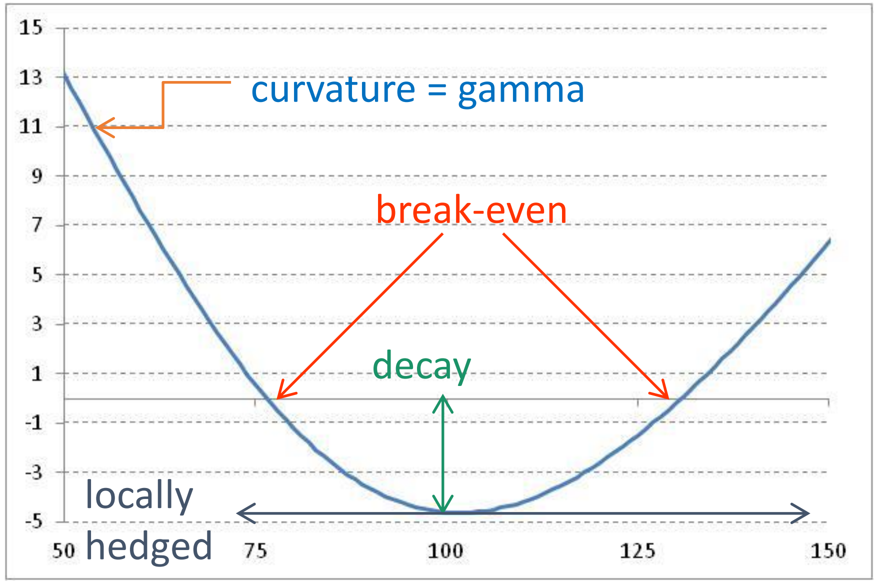
- Black swans likely missing from historical data
  - Example: 3y S&P volatility estimate 2015-2016-2017 = 12.50%
  - What to expect in 2018?
  - Assume probability of major sell-off = 3% (frequency of global crises over 100 years)
  - In this case we may hit 2008 volatility = 40%
  - Otherwise expect average of 12.50%
  - (reasonable but very simplified view, of course)
  - Average expected volatility = 3% * 40% + 97% * 12.50% = 13.50%
  - Possibility of black swan adds 1 volatility point over historical estimate
  - Interestingly, 1Y ATM implied = 13.50% as of end Dec 2017

- Historical volatility estimates should be adjusted for possibility of black swans
- By definition, likely missing from historical data
- But incorporated in forward looking implied volatility

Hedging a European option in Black-Scholes

- Simplified context
  - No rates, no dividends, no credit, zero-cost short selling

- Buy a call option maturity T strike K for implied volatility σ
- This means the price is Ct=BS(St,t;σ)
- How does option price move when spot moves?
- BS sensitivities:
  - 1st order sensitivity to S = delta
  - 2nd order sensitivity to S = gamma > 0
  - Sensitivity to time = theta < 0
- By Ito: $\underbrace{dC_t = dBS\left(S_t, t; \sigma\right)}_{\text{change in option price}} = \underbrace{\vartheta dt}_{\text{decay}} + \underbrace{\Delta S_t \underbrace{\left(\frac{dS_t}{S_t}\right)}_{\text{actual daily return}}}_{\text{delta impact}} + \underbrace{\frac{1}{2}\Gamma S_t^2 \underbrace{\left(\frac{dS_t}{S_t}\right)^2}_{\text{sqaure of actual daily return}}}_{\text{gamma impact}}$
- Hedging delta (selling delta stocks) neutralizes delta:
  - Hence daily PnL is: $dC_t - \Delta S_t = \vartheta dt + {\color{red}0} + \frac{1}{2}\Gamma\left(S_t\right)^2\left(\frac{dS_t}{S_t}\right)^2$



- Being long an option and hedging
  - We lose decay if spot does not move
  - And compensate with positive PnL
    = quadratic function of **realized** spot return

- What is the break-even spot return?

The Fundamental Theorem of Derivatives Trading

- Denote the **realized** volatility $\sigma_r \equiv \sqrt{\left(\frac{dS}{S}\right)^2 \Big/ dt}$
- Remember the PDE satisfied by the Black-Scholes formula (historically, it was produced by solving that PDE): $\theta + \frac{1}{2}\Gamma S^2 \sigma^2 = 0$
- Injecting into the PnL equation: $\boxed{PnL = \frac{1}{2}\Gamma S_t^2 \left(\sigma_r^2 - \sigma^2\right) dt}$
- Delta-hedging options = **swapping implied for realized variance**
- In particular PnL break-even is one implied standard deviation away: $PnL = 0 \Leftrightarrow \sigma_r^2 = \sigma^2 \Leftrightarrow \frac{dS}{S} = \pm\sigma\sqrt{dt}$
- Note the proportion coefficient $\frac{1}{2}\Gamma S_t^2$: we can choose strikes in a long portfolio of options to make it constant
- Then the final PnL of hedging all the way to expiry is proportional to realized – implied variance
- Such portfolio is called a log-contract because $\Gamma S_t^2 = k \Leftrightarrow \Gamma = k/S_t^2 \Leftrightarrow \Delta = k/S_t + k_2 \Leftrightarrow C = \log\left(S_t/k_3\right)$
- Note the VIX is computed as the price of a log-contract
- And variance swaps are valued and hedged this way

Take-Away: Robustness of Black-Scholes (and other arbitrage-free models)

- Implied volatility is **not** some arbitrary unit for measuring prices

- A simple strategy **monetizes** implied volatility: hedging options exchanges realized for implied volatility

- Many volatility hedge funds and options traders (called gamma players) constantly look for mispriced implied volatility
- In the sense that it does not correspond to the predicted future realized
- And realize the arbitrage until exhaustion
- Hence the implied volatility **must** correspond to the market's best estimate of future realized

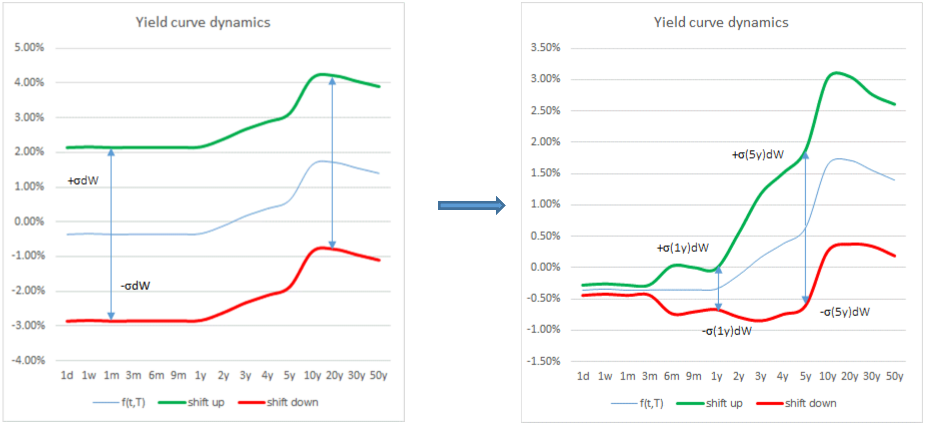
- We normally **calibrate** our models to reflect implied volatility
  - Because we must hedge the volatility risk of complex options with Europeans
  - Hence Europeans are hedge instruments, and the model must match their value

- That must be identical to using correctly estimated historical volatility
  - Because implied = best predictor from a large number of professional traders
  - Willing to wage monetary stakes on that predictor

Heath-Jarrow-Morton models

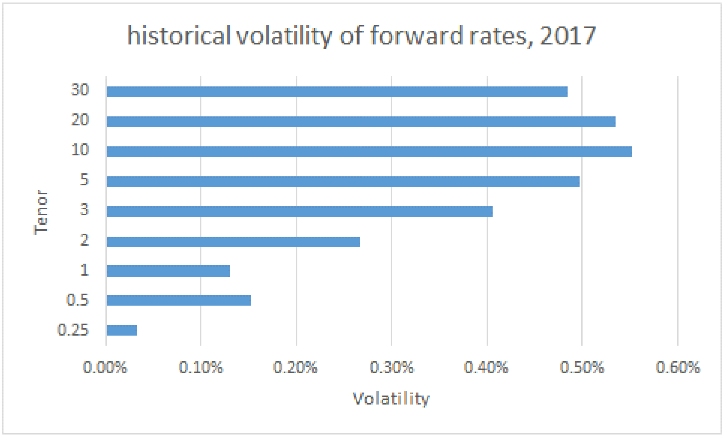
HJM (1992)

One-factor Gaussian HJM model (1992)



- An extension of our simplistic parallel yield curve model $df(t,T) = \sigma dW$
- Where we allow more general, non parallel, deformations: $df(t,T) = \sigma(t,T)dW$
  - At time t, $T \to \sigma(t,T)$ is the deformation of the YC between t and t+dt
  - Future deformations are allowed to change with time t
  - Special case: stationary model where deformations are **constant in tenor** = remaining maturity $df(t,T) = \sigma(T-t)dW$
- Rates are still Gaussian (sigma deterministic) and 100% correlated (one factor) but no longer same volatility (flat deformations)
- Historical specification: stationary model, directly estimate $\sigma(T-t)$
- Implied calibration: set the 2D surface $\sigma(t,T)$

  to fit today's 2D surface of ATM swaption prices



HJM dynamics

- By the exact same arguments as in the parallel model = convexity arbitrage
  - It must exist a unique risk premium process η
  - Such that the dynamics of forward rates under the historical probability is $$\boxed{df(t,T) = \left[\underbrace{\sigma(t,T)\int_t^T \sigma(t,u)\,du}_{\text{convexity arbitrage adjustment}} - \underbrace{\eta_t \sigma(t,T)}_{\text{risk premium}}\right] dt + \underbrace{\sigma(t,T)\,dW}_{\text{random shifts}}}$$
  - The dynamics of all asset prices A, bonds and derivatives, is:

$$\frac{dA_t}{A_t} = \left[\underbrace{r_t}_{\text{short rate} = f(t,t)} + \underbrace{\eta_t \sigma^A_t}_{\text{risk premium}}\right] dt + \underbrace{\sigma^A_t}_{\text{volatility of asset}} dW$$

  - In particular, for bond prices, we have:

$$\sigma_{DF}(t,T) = -\int_t^T \sigma(t,u)\,du \text{ hence } \frac{dDF(t,T)}{DF(t,T)} = \left[r_t + \eta_t \int_t^T \sigma(t,u)\,du\right] dt - \left[\int_t^T \sigma(t,u)\,du\right] dW$$

- It follows that the prices of all primary and derivatives assets satisfy, under the historical probability:

$$A_t = E_t\left\{\exp\left[-\int_t^T \left[r_s + \eta_s \sigma^A_s\right] ds\right] A_T\right\} \text{ in particular } A_0 = E\left\{\exp\left[-\int_0^{T^*} \left[r_s + \eta_s \sigma^A_s\right] ds\right] A_{T^*}\right\},\ A_{T^*} = payoff$$

Prices are expectations, discounted at a rate adjusted for the risk (volatility) of each asset

HJM risk neutralization

- As in the parallel model, prices are independent of risk premium
  - Discounting and drift compensate
  - We obtain the same results with 0 risk premium
  - Using the risk-neutral rate dynamics: $\boxed{df(t,T) = \left[\sigma(t,T)\int_t^T \sigma(t,u)\,du\right]dt + \sigma(t,T)\,dW}$ (Hea th-Jarrow-Morton formula)
  - And we have the risk-neutral pricing equation:

$$A_t = E_t\left\{\exp\left[-\int_t^T r_s ds\right]A_T\right\} \text{ in particular } A_0 = E\left\{\exp\left[-\int_0^{T*} r_s ds\right]A_{T*}\right\}, A_{T*} = payoff$$

- We generalized almost all results from the parallel model
- Except the Markov property
- It turns out that HJM is **not** Markov : the short rate is **not** a self contained diffusion under the risk-neutral probability

HJM: practical implementation

- No Markov property: the state vector at time t is <u>the entire yield curve</u>
  - Large (infinite) dimension: finite difference impractical (cost exponential in dimension)
  - Monte-Carlo simulation slow: keep track of entire curve at every time step
  - Must somehow discretize continuous curve and dynamics
- Subject to overfitting
  - Large number of parameters
  - May sample unrealistic future curves
- No fast (analytical, FDM) pricing of European options for calibration

➔ HJM is more a theoretical framework than a practical model

- Libor Market Models (Brace-Gatarek-Musiela, 1995) offer a practical solution
  - Discrete HJM, model forward libors instead of instantaneous forward rates
  - Exact formula for caps, approximate formula for swaptions
  - But still subject to overfitting and costly Monte-Carlo simulations
- Practical choice:
  1. Implement LMM over data centres/GPU
  2. Or restrict to Markov specification
  - We will see that multi-factor Markov models offer same flexibility
  - And substantially higher speed ➔ suitable for (multi-threaded) CPU implementation

Markov HJM models

Cheyette (1992)

Markov HJM: fundamental result

- We remind that a model is Markov when the short rate is a self contained diffusion

- Fundamental result
  - Independently derived by a number of researchers in the early to mid 1990s
  - Most general form derived by Cheyette in 1992 ➔ the name “Cheyette model” stuck

- The 1F Gaussian HJM is Markov (r is a self contained diffusion) if (and only if) volatility is **separable**

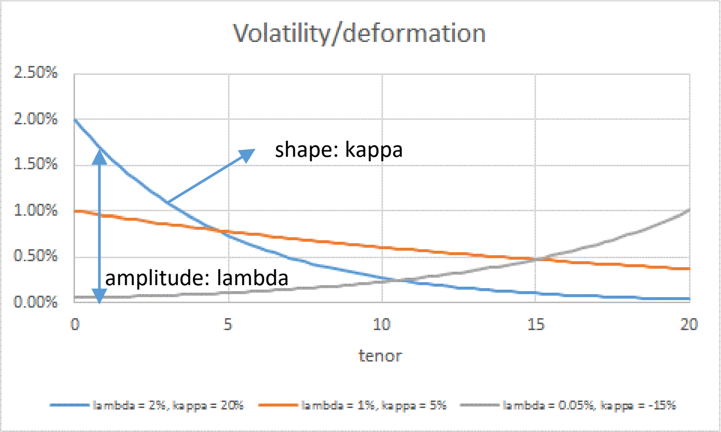
$$\boxed{\sigma(t,T)=\lambda(t)\exp\left(-\int_t^T k(u)\,du\right)}$$

- Restricted class of HJM models
  - Rate volatility not a 2D surface, but 2 curves λ and κ
  - κ controls the shape of the deformations
  - λ represents the amplitude of the deformations

- Customary (but not compulsory) choice: deformations of stationary shape

  - Constant kappa: $\sigma(t,T)=\lambda(t)\exp\left[-k(T-t)\right]$

  - Rate volatility / YC deformations exponentially decreasing in tenor
  - In what follows, we systematically use constant kappa to simplify equations



Markov HJM: short rate dynamics

- It can be shown (with a bit of calculus) that under the **risk-neutral** probability:

$$dr_t = \left\{ \underbrace{\frac{\partial f(0,t)}{\partial t}}_{\text{follows forwards}} - \underbrace{k\left[r_t - f(0,t)\right]}_{\substack{\text{kappa is the mean-reversion} \\ \text{of the short rate towards forwards}}} + \underbrace{y(t)}_{\text{convexity adjustment}} \right\} dt + \underbrace{\lambda(t)}_{\substack{\text{lambda is the volatility} \\ \text{of the short rate}}} dW$$

$$y(t) = \int_0^t \lambda^2(u) \exp\left[-2k(t-u)\right] du \text{ : deterministic}$$

(Historical probability: subtract risk premium times lambda in the drift)

- Reconstruction formula: $f(t,T) = \underbrace{f(0,T)}_{\text{today's forward}} + \exp\left[-k(T-t)\right] \underbrace{\left[r_t - f(0,t)\right]}_{\text{factor}} + \underbrace{\frac{\exp\left[-k(T-t)\right] - \exp\left[-2k(T-t)\right]}{k} y(t)}_{\text{convexity adjustment}}$

- We define the factor $X_t \equiv r_t - f(0,t)$
  - Then: $dX_t = \left\{-k\left[r_t - f(0,t)\right] + y(t)\right\} dt + \lambda(t) dW$
  - And we have the reconstruction formula: $f(t,T) = f(0,T) + \exp\left[-k(T-t)\right] X_t + \frac{\exp\left[-k(T-t)\right] - \exp\left[-2k(T-t)\right]}{k} y(t)$

Markov HJM: benefits

- Clear, explicit, exponential deformations

➔ Intuition/control over future YC shape

- Short rate = self contained diffusion
  + closed form reconstruction formula

➔ Simulate whole curve with only one state variable

  - Ultra fast finite difference in dimension 1
  - Fast Monte-Carlo simulations

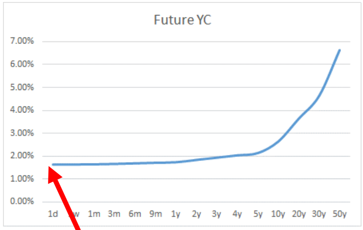
- Closed form formulas for swaption prices
  see e.g. Jamshidian, 1989

➔ Fast calibration

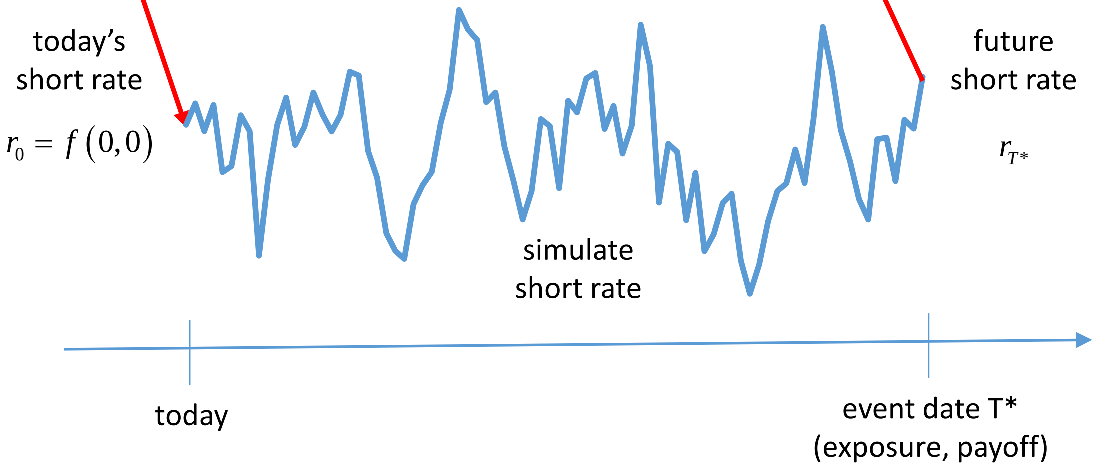
reconstruct future YC

$$f(t,T) = f(0,T) + \exp\left[-k(T-t)\right]\left[r_t - f(0,t)\right] + \frac{\exp\left[-k(T-t)\right] - \exp\left[-2k(T-t)\right]}{k} y(t)$$

$f(0,t)$



$$dr_t = \left\{\frac{\partial f(0,t)}{\partial t} - k\left[r_t - f(0,t)\right] + y(t)\right\} dt + \lambda(t)\, dW$$



Brief introduction to Cheyette (1992) model

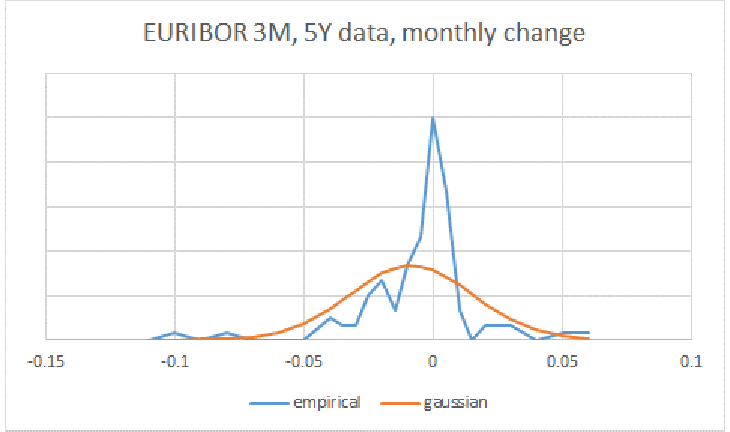
- 1F Gaussian Markov HJM: $dX_t = \{-k[r_t - f(0,t)] + y(t)\}dt + \lambda(t)dW$

$$y(t) = \int_0^t \lambda^2(u)\exp[-2k(t-u)]du \text{ : deterministic}$$

$$f(t,T) = f(0,T) + \exp[-k(T-t)]X_t + \frac{\exp[-k(T-t)] - \exp[-2k(T-t)]}{k}y(t)$$

All rates have Gaussian distribution

- Not verified in practice
  - Skew and kurtosis in historical data
  - Implied volatility “smile” in swaption data



Cheyette (1992)

- 1F Gaussian Markov HJM: lambda = amplitude of deformations = **deterministic** function of time

- Cheyette (1992): lambda = stochastic process
  - May depend on short rate = local volatility
  - May be stochastic on its own right = stochastic volatility
  - Or a mix of both to fit empirical distributions or market smiles

$$dX_t = \left[-kX_t + y(t)\right]dt + \lambda\left(t, X_t, W^{\lambda}\right)dW$$

$$y(t) = \int_0^t \lambda^2\left(u, X_u, W^{\lambda}{}_u\right)\exp\left[-2k(t-u)\right]du \text{ : no longer deterministic}$$

$$f(t,T) = f(0,T) + \exp\left[-k(T-t)\right]X_t + \frac{\exp\left[-k(T-t)\right] - \exp\left[-2k(T-t)\right]}{k}y(t) \text{ : holds identical}$$

- Cheyette's result:
  - All of the previous properties hold, in particular reconstruction formula is identical
  - Note y is no longer deterministic: Markov dimension is 2 ➔ technical adjustments in FDM and Monte-Carlo
  - Closed-form results for swaptions are also lost, but approximations exist, see e.g. Andreasen, 2005

Markov HJM: under parameterization

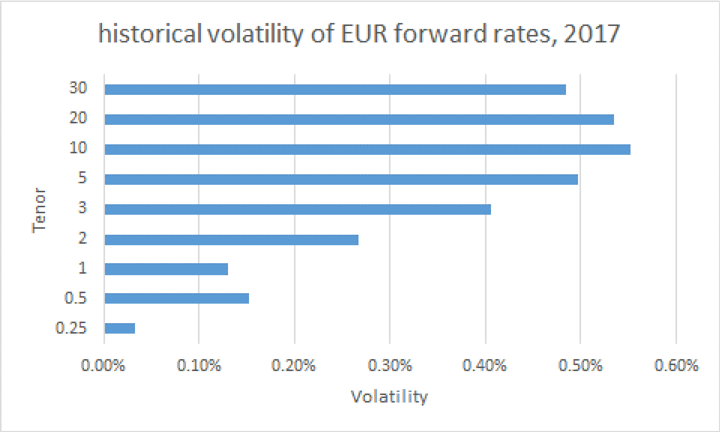
- Historical estimation of forward rate volatility:
  - Short tenors: very little volatility due to central bank action
  - Volatility increases with tenor up to 7-10y
  - Then decreases (due to mean-reversion?)

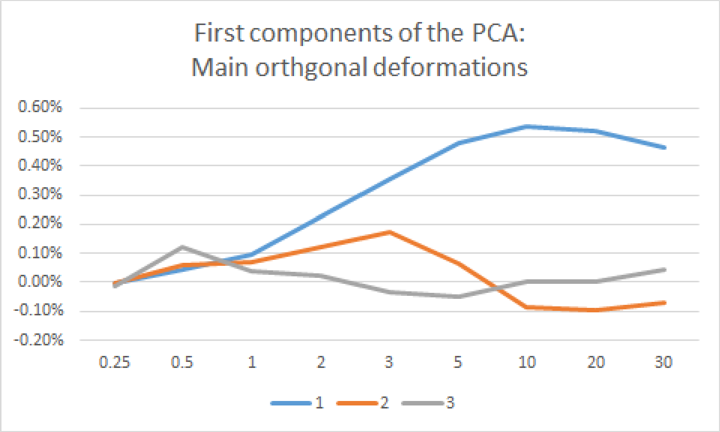
➔ Main deformation (1st component of historical PCA)
Same shape, increases up to ~10y, then decreases

- The same structure implied volatility = f(tenor)
  seen in volatility implied from swaption prices
- Stationary Markov HJM:
  only exponentially increasing/decreasing deformations

➔ Cannot fit realistic dynamics or implied volatility

- We only have one kappa and a curve lambda (t)
  - Historical calibration: set lambda to "average" tenor volatility
  - Implied calibration: calibrate lambda(t) to one tenor per expiry or best fit all tenors for each expiry
  - In all cases, (severely) miss volatility of (most) tenors
- 1F Markov HJM is under parameterized
  and cannot properly fit realistic dynamics or market
- These problems are (neatly) resolved with multi-factor models

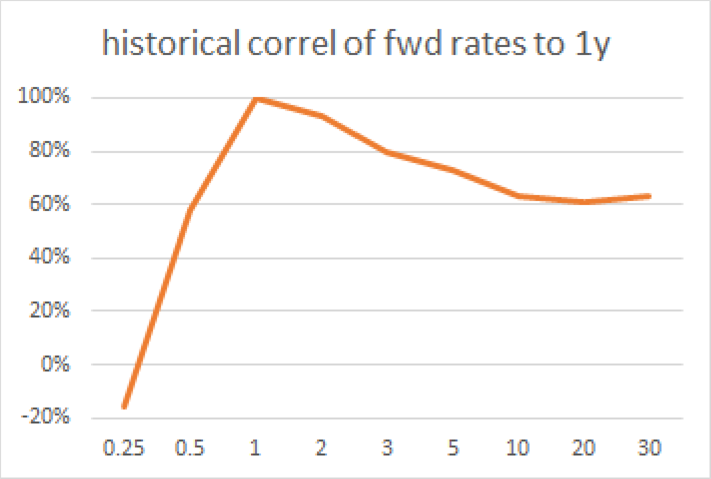




Multi-factor models

From HJM (1992) to MFC (Andreasen, 2005)

Multi-factor interest rate models



- General 1F HJM can fit arbitrary rate volatility
  (But Markov specification can only fit very restricted, unrealistic volatility structures per tenor)
- But in 1F models correlation between rates is always 100%
- That may be a problem for the risk management of exotics
- And a **serious** problem for exposure estimation
  - In 1F models, a 10y receiver swap may be perfectly hedged with a 2y payer swap
  - In reality, this is not the case: 2017 historical correlation between 2y and 10y forward EUR libors is only 75%
  - ➔ 1F models may overestimate netting effects within swap portfolios
    And (severely) underestimate exposures
- More generally
  - Exposures are mainly "basket options" on trades in a netting set
  - And it is well known that these options are highly sensitive on correlation
- We need **multi-factor** models
  - Produce correct correlation between rates of different maturities
  - As an additional benefit, resolve volatility fitting in Markov models

Two-factor Gaussian HJM

- HJM, like Cheyette, was initially written as a multi-factor model

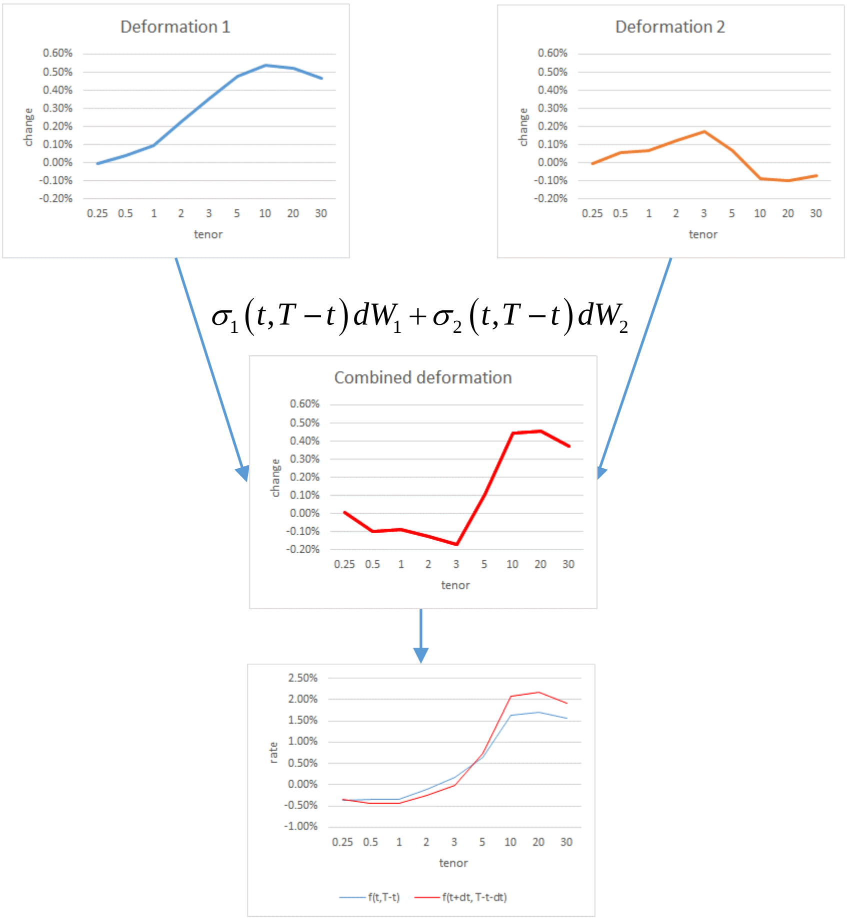
- For pedagogical reasons only,
  - We started with a parallel (kappa = 0) 1F model
  - Generalized into 1F HJM / Markov HJM / Cheyette
  - New review 2F specification
  - And finally generalize to n factors (typically 4 factors)
    ➔ Trade off between speed and fitting of covariance structure across tenors

- Two-factor HJM model
  - Two simultaneous deformations
  - Each driven by (possibly correlated) Brownian motions
  - Producing the effective deformation of the YC

- Historical calibration: first 2 components of the PCA, uncorrelated Brownians
- Implied calibration: match historical correlation and swaption prices
  - Out of scope for this talk --- we focus on Markov models
  - See for instance, Andersen-Piterbarg, volume 3

$\sigma_1(t, T-t)$ $\sigma_2(t, T-t)$

$$\sigma_1(t, T-t)\,dW_1 + \sigma_2(t, T-t)\,dW_2$$



$$f(t+dt, T-t-dt) = f(t, T-t) + \sigma_1(t, T-t)\,dW_1 + \sigma_2(t, T-t)\,dW_2$$

2F HJM: arbitrage-free dynamics

- 2F HJM $df(t,T) = \ldots + \sigma_1(t,T)dW_1 + \sigma_2(t,T)dW_2$, $\langle dW_1, dW_2 \rangle = \rho dt$

- As before, we have the convexity arbitrage
  - Buy long term bond (e.g. 10y)
  - Hedge its **2 risks with 2 short term bonds** (e.g. 1y and 3y)
  - Enjoy free convexity

- Hence, no arbitrage dictates that
  - It must exist **2 unique risk premiums** (one per factor) η_t^1, η_t^2
  - Such that all asset prices (and in particular bond prices) satisfy

$$\frac{dA_t}{A_t} = \left[r_t + \eta_t^1 \sigma^{A1}{}_t + \eta_t^2 \sigma^{A2}{}_t\right]dt + \sigma^{A1}{}_t dW_1 + \sigma^{A2}{}_t dW_2 \quad \text{---} \quad \frac{dDF(t,T)}{DF(t,T)} = \left[r_t + \eta_t^1 \int_t^T \sigma_1(t,u)du + \eta_t^2 \int_t^T \sigma_2(t,u)du\right]dt - \left[\int_t^T \sigma_1(t,u)du\right]dW_1 - \left[\int_t^T \sigma_2(t,u)du\right]dW_2$$

  - Which implies that prices are expectations of payoffs discounted with risk-adjusted rates, and the forward rate dynamics:

$$df(t,T) = \left[\sigma_1(t,T)\int_t^T \sigma_1(t,u)du + \sigma_2(t,T)\int_t^T \sigma_2(t,u)du + \rho\sigma_1(t,T)\int_t^T \sigma_2(t,u)du + \rho\sigma_2(t,T)\int_t^T \sigma_1(t,u)du - \eta_t^1\sigma_1(t,T) - \eta_t^2\sigma_2(t,T)\right]dt + \sigma_1(t,T)dW_1 + \sigma_2(t,T)dW_2$$

- What is interesting here is we now have 2 risk premium, one per Brownian / factor
- Asset prices don't depend on risk premium, we can price assets as short rate discounted expectations under the RN dynamics:

$$\boxed{df(t,T) = \left[\sigma_1(t,T)\int_t^T \sigma_1(t,u)du + \sigma_2(t,T)\int_t^T \sigma_2(t,u)du + \rho\sigma_1(t,T)\int_t^T \sigma_2(t,u)du + \rho\sigma_2(t,T)\int_t^T \sigma_1(t,u)du\right]dt + \sigma_1(t,T)dW_1 + \sigma_2(t,T)dW_2}$$

2F Markov HJM = 2F Cheyette

- As a direct extension of the 1F case
  The 2F Gaussian HJM model is Markov if and only if both volatility functions are separable:

$$\sigma_1(t,T)=\lambda_1(t)\exp\left(-\int_t^T k_1(u)\,du\right) \quad , \quad \sigma_2(t,T)=\lambda_2(t)\exp\left(-\int_t^T k_2(u)\,du\right)$$

- In what follows, we only consider stationary shape specifications (constant kappas): $\sigma_i(t,T)=\lambda_i(t)\exp\left[-k_i(T-t)\right]$

- We have the 2F dynamics:
  - Gaussian case:
    - y's are deterministic
    - Markov dimension is 2: X1 and X2
  - Local / stochastic lambda case
    - y's are path-dependents
    - Markov dimension is 5

$$dX_1=\left\{-k_1X_1+y_{11}(t)+y_{12}(t)\right\}dt+\lambda_1(t)\,dW_1$$

$$dX_2=\left\{-k_2X_2+y_{22}(t)+y_{12}(t)\right\}dt+\lambda_2(t)\,dW_2$$

$$y_{11}(t)=\int_0^t\lambda_1^2(u)\exp\left[-2k_1(t-u)\right]du$$

$$y_{22}(t)=\int_0^t\lambda_2^2(u)\exp\left[-2k_2(t-u)\right]du$$

$$y_{12}(t)=\int_0^t\rho\lambda_1(u)\lambda_2(u)\exp\left[-(k_1+k_2)(t-u)\right]du$$

- We have the reconstruction formula:

$$\begin{aligned}f(t,T)=\;&f(0,T)+\exp\left[-k_1(T-t)\right]X_1+\exp\left[-k_2(T-t)\right]X_2\\&+\frac{\exp\left[-k_1(T-t)\right]-\exp\left[-2k_1(T-t)\right]}{k_1}y_{11}(t)+\frac{\exp\left[-k_2(T-t)\right]-\exp\left[-2k_2(T-t)\right]}{k_2}y_{22}(t)\\&+\frac{k_1\exp\left[-k_1(T-t)\right]+k_2\exp\left[-k_2(T-t)\right]-(k_1+k_2)\exp\left[-(k_1+k_2)(T-t)\right]}{k_1k_2}y_{12}(t)\end{aligned}$$

- And closed-form formulas for swaptions:
  - Gaussian case: exact
  - Local / stochastic lambda case: approximate, see Andreasen, 2005

2F Cheyette: historical calibration

- The stationary 2F Cheyette model: $df(t,T) = \lambda_1 \exp\left[-k_1(T-t)\right]dW_1 + \lambda_2 \exp\left[-k_2(T-t)\right]dW_2 + \mu_f(t,T)dt \; , \; \langle dW_1, dW_2 \rangle = \rho dt$

$$\mu_f(t,T) = \frac{\rho\lambda_1\lambda_2}{k_1k_2}\left\{k_1 \exp\left[-k_1(T-t)\right] + k_2 \exp\left[-k_2(T-t)\right] - (k_1+k_2)\exp\left[-(k_1+k_2)(T-t)\right]\right\}$$

- Is parameterized by 5 parameters:
    - The 2 (exponential) deformation shapes kappa 1 and kappa 2
    - The amplitudes of the 2 deformations lambda 1 and lambda 2
    - The correlation rho between the 2 factors

- It generates the following covariance structure between forward rates of different tenors:

$$\frac{\operatorname{cov}\left[df(t,t+\tau_1), df(t,t+\tau_2)\right]}{dt} = \lambda_1^2 \exp\left[-k_1(\tau_1+\tau_2)\right] + \lambda_2^2 \exp\left[-k_2(\tau_1+\tau_2)\right] + \rho\lambda_1\lambda_2\left[\exp(-k_1\tau_1 - k_2\tau_2) + \exp(-k_2\tau_1 - k_1\tau_2)\right]$$

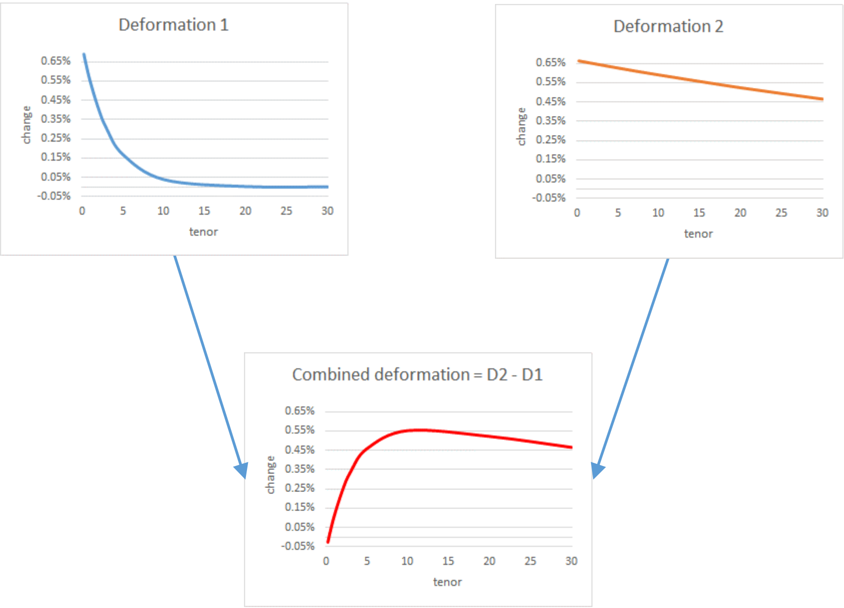
- We can best fit the historical covariance matrix in the 5 parameters

- For EUR rates over 2017, we obtain:
    - kappa 1 = 30%, kappa 2 = 1%
    - lambda 1 = 1%, lambda 2 = 0.65%
    - Rho -99%

2F Cheyette: historical calibration results

Calibrated model

$\lambda_1 \exp(-k_1 \tau)$ $\qquad\qquad$ $\lambda_2 \exp(-k_2 \tau)$

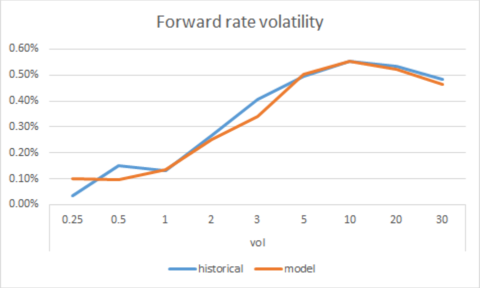


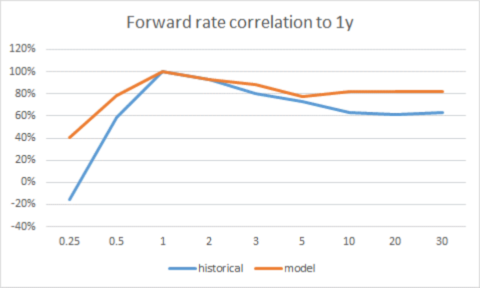
$$\rho \approx -100\%$$

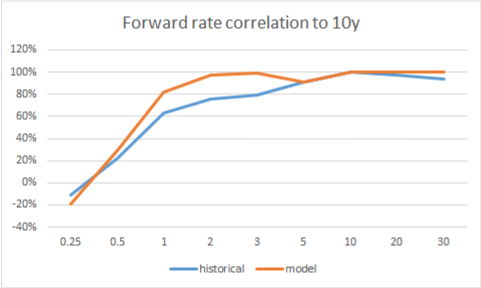
$$\Rightarrow$$

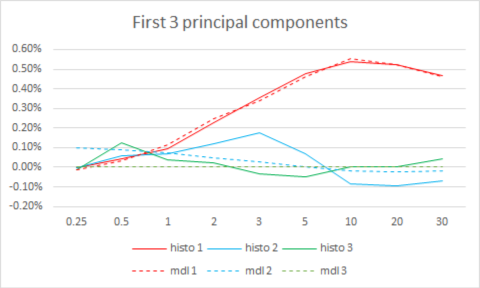
$$df(t, t+\tau) \approx \left[\lambda_1 \exp(-k_1 \tau) - \lambda_2 \exp(-k_2 \tau)\right] dW$$

Results









2F Cheyette: towards a mixed calibration

- The 1F Cheyette could not produce realistic deformations

- The 2F extension clearly solved that flaw

- In addition, we are able to produce a (reasonably) sensible correlation structure

- But calibration to implied data (swaption prices) is unclear
  - We typically want to calibrate to **historical** correlations
  - But **imply** volatility from swaption prices
  - In other terms, set correlation of rates historically, and fit volatility to swaption prices
  - It is unclear how to do that since the model is not parameterized in terms of forward rate volatility and correlation
  - But shape (kappas), amplitude (lambdas) and correlation (rho) of factors

- Our preferred solution is explained next

2F Cheyette: re-basing (Andreasen)

- We now allow a time-dependent lambda and rho: $df(t,T) = \lambda_1(t)\exp\left[-k_1(T-t)\right]dW_1 + \lambda_2(t)\exp\left[-k_2(T-t)\right]dW_2 + \mu_f(t,T)dt \ , \ \langle dW_1, dW_2 \rangle = \rho(t)dt$

- We pick **2 skew tenors** (because the model is 2F) $\tau 1$ and $\tau 2$, for instance 1y and 10y

- We arbitrarily set the 2 kappas to fixed values

- At time t, we have an expression for the volatility and correlation between running rates of tenors $\tau 1$ and $\tau 2$

$$\sigma_{f_1}(t)^2 = \lambda_1(t)^2 \exp(-2k_1\tau_1) + \lambda_2(t)^2 \exp(-2k_2\tau_1) + 2\rho(t)\lambda_1(t)\lambda_2(t)\exp\left[-(k_1+k_2)\tau_1\right]$$
$$\sigma_{f_2}(t)^2 = \lambda_1(t)^2 \exp(-2k_1\tau_2) + \lambda_2(t)^2 \exp(-2k_2\tau_2) + 2\rho(t)\lambda_1(t)\lambda_2(t)\exp\left[-(k_1+k_2)\tau_2\right]$$
$$\rho_{f_1,f_2}\sigma_{f_1}(t)\sigma_{f_2}(t) = \lambda_1(t)^2 \exp\left[-k_1(\tau_1+\tau_2)\right] + \lambda_2(t)^2 \exp\left[-k_2(\tau_1+\tau_2)\right] + \rho(t)\lambda_1(t)\lambda_2(t)\left[\exp(-k_1\tau_1 - k_2\tau_2) + \exp(-k_2\tau_1 - k_1\tau_2)\right]$$

- We invert that formula: solve for lambdas and rho **given** the volatility and correlation of the skew rates

$$\begin{bmatrix} \lambda_1^2 \\ \lambda_2^2 \\ \rho\lambda_1\lambda_2 \end{bmatrix}(t) = \begin{bmatrix} \exp(-2k_1\tau_1) & \exp(-2k_2\tau_1) & 2\exp\left[-(k_1+k_2)\tau_1\right] \\ \exp(-2k_1\tau_2) & \exp(-2k_2\tau_2) & 2\exp\left[-(k_1+k_2)\tau_2\right] \\ \exp\left[-k_1(\tau_1+\tau_2)\right] & \exp\left[-k_2(\tau_1+\tau_2)\right] & \exp(-k_1\tau_1 - k_2\tau_2) + \exp(-k_2\tau_1 - k_1\tau_2) \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{f_1}^2 \\ \sigma_{f_2}^2 \\ \rho_{f_1,f_2}\sigma_{f_1}\sigma_{f_2} \end{bmatrix}(t)$$

2F Cheyette: re-basing (2)

- This allows us to parameterize the model in terms of volatility and correlation of the skew rates
    - Historical calibration ➔ set to historical volatilities and correlations
    - Implied calibration:
        - Set correlation to historical
        - Leave (time-dependent) volatilities as degrees of freedom to calibrate 2 ATM swaptions of 2 different tenors by expiry

- Which internally resolves into 2 lambdas and 1 rho with the matrix inversion

- That fixes the volatilities and correlation structure of other rates (by interpolation/extrapolation, dependent on kappas)

- Hence, kappas act as interpolation / extrapolation parameters
  From skew tenors, which volatility and correlation are set
  To other tenors

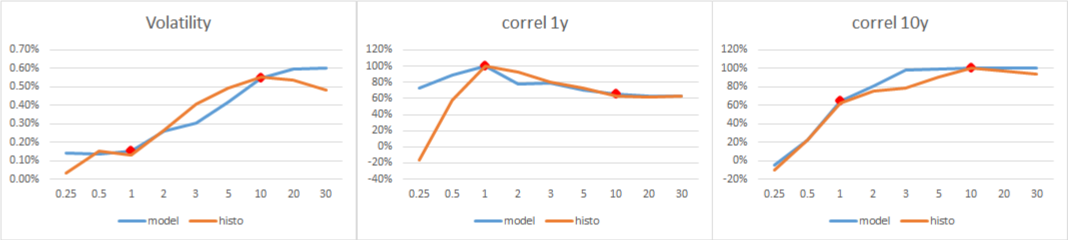
- This means that with more factors (and hence skew tenors) kappas have little impact

2F Cheyette: re-basing example

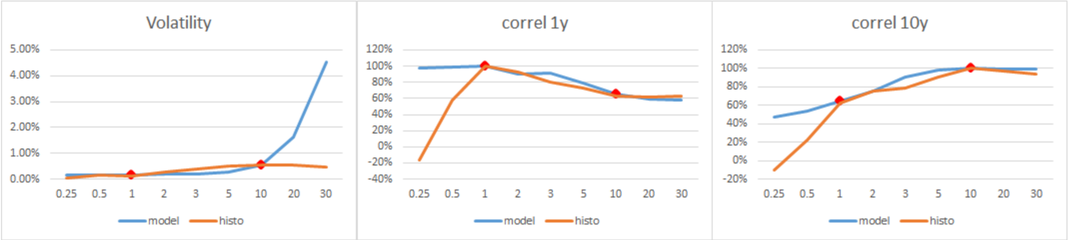
- We use 1y and 10y as skew rates

- With a historical estimation over 2017:
  - Volatility of 1y tenor = 15 bppa
  - Implied of 10y tenor = 55 bppa
  - Correlation (1y, 10y) = 65%

- With kappa 1 = 0.01% (flat)
  and kappa 2 = 25% (rapidly decreasing) we get:
  - lambda 1 = 0.60%
  - lambda 2 = 0.67%
  - rho = -97%

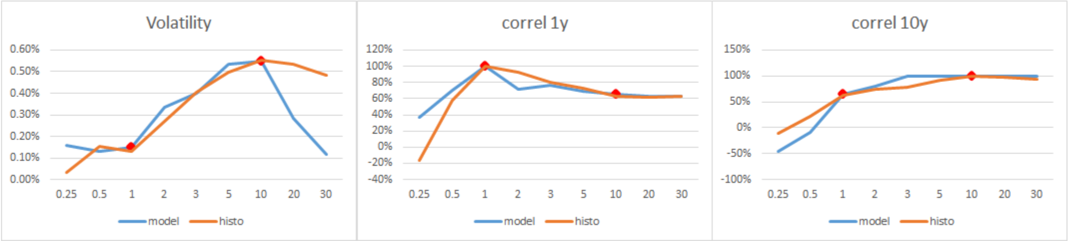


- With kappa 1 = -10% (increasing)
  and kappa 2 = 10% (decreasing):

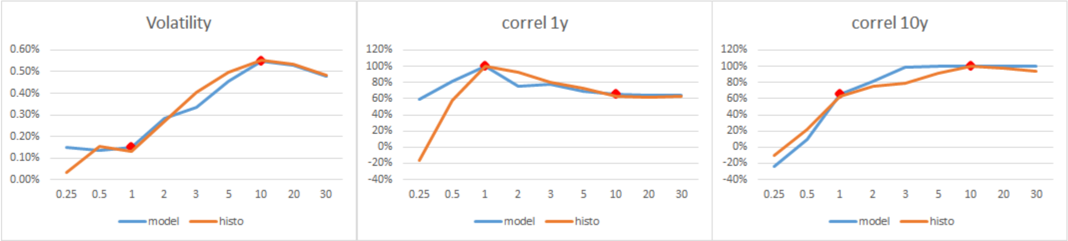


(obviously not a good if we extrapolate to long maturities)

- kappa 1 = 10% and kappa 2 = 20%



- kappa 1 = 30% and kappa 2 = 1%



2F Cheyette: re-basing example (2)

- Unsurprisingly, interpolation is reasonably stable but not extrapolation

➔ Best practice: use shortest and longest rate as skew tenors, + a few in between

➔ Two factors is not enough!

- We typically need 4 factors with 4 skew tenors: 6m, 2y, 10y, 30y
- We review n-factor extension next

N factor HJM

- N simultaneous deformations:

$$df\left(t,T\right)=\sum_{i=1}^{n}\sigma_i\left(t,T\right)dW_i\ ,\ \left\langle dW_i,dW_j\right\rangle=\rho_{ij}dt$$

- Arbitrage-free dynamics: n (unique) risk premiums

  - Historical probability:
    All asset prices are expectations of future prices
    discounted with risk adjusted rate

$$\frac{dA_t}{A_t}=\left(r_t+\sum_{i=1}^{n}\eta_t^i\sigma^{Ai}{}_t\right)dt+\sum_{i=1}^{n}\sigma^{Ai}{}_t dW_i$$

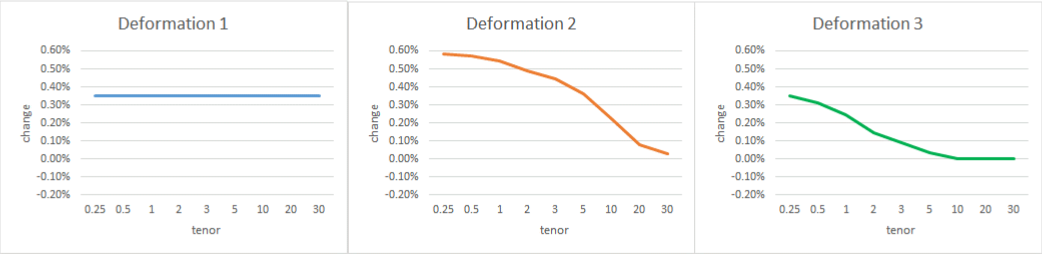
$$\frac{dDF\left(t,T\right)}{DF\left(t,T\right)}=\left[r_t+\sum_{i=1}^{n}\eta_t^i\int_t^T\sigma_i\left(t,u\right)du\right]dt-\sum_{i=1}^{n}\left[\int_t^T\sigma_i\left(t,u\right)du\right]dW_i$$

$$df\left(t,T\right)=\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}\sigma_i\left(t,T\right)\int_t^T\sigma_j\left(t,u\right)du-\sum_{i=1}^{n}\eta_t^i\sigma^{Ai}{}_t\right]dt+\sum_{i=1}^{n}\sigma_i\left(t,T\right)dW_i$$

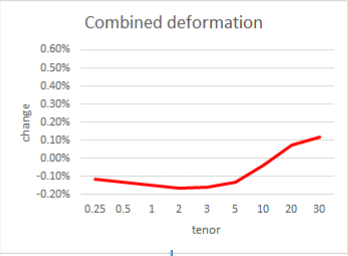
  - Risk neutral probability
    - Prices are independent of risk premiums ➔ we can set them to 0
    - Asset prices are expectations of future prices
      discounted with risk free rate
    - Rate dynamics under the risk neutral probability:

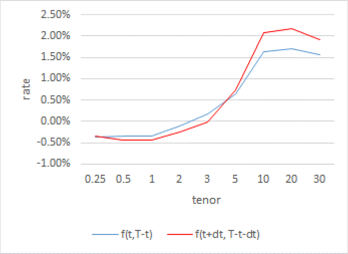
$$df\left(t,T\right)=\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}\sigma_i\left(t,T\right)\int_t^T\sigma_j\left(t,u\right)du\right]dt+\sum_{i=1}^{n}\sigma_i\left(t,T\right)dW_i$$

$$\sigma_i\left(t,T-t\right)$$



$$\sum_{i=1}^{n}\sigma_i\left(t,T-t\right)dW_i$$





$$f\left(t+dt,T-t-dt\right)=f\left(t,T-t\right)+\sum_{i=1}^{n}\sigma_i\left(t,T-t\right)dW_i$$

N factor HJM: special cases

- Stationary:

$$\sigma_i(t,T) = \sigma_i(T-t)$$

- Markov:

$$\sigma_i(t,T) = \lambda_i(t)\exp\left[-\int_t^T k_i(u)\,du\right]$$

- Stationary shape Markov:

$$k_i = cste \Rightarrow \sigma_i(t,T) = \lambda_i(t)\exp\left[-k_i(T-t)\right]$$

- Factor dynamics:

$$dX_i = \left\{-k_i X_i + \sum_{j=1}^{n} y_{ij}(t)\right\}dt + \lambda_i(t)\,dW_i$$

$$y_{ij}(t) = \int_0^t \rho_{ij}\lambda_1(u)\lambda_2(u)\exp\left[-\left(k_i+k_j\right)(t-u)\right]du$$ Gaussian: y's deterministic , otherwise: y's path-dependent

- Reconstruction formula:

$$f(t,T) = f(0,T) + \sum_{i=1}^{n}\exp\left[-k_i(T-t)\right]X_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{k_i\exp\left[-k_i(T-t)\right] + k_j\exp\left[-k_j(T-t)\right] - \left(k_i+k_j\right)\exp\left[-\left(k_i+k_j\right)(T-t)\right]}{k_i k_j}\,y_{ij}(t)$$

- Closed-form approximations for swaptions: see Andreasen, 2005

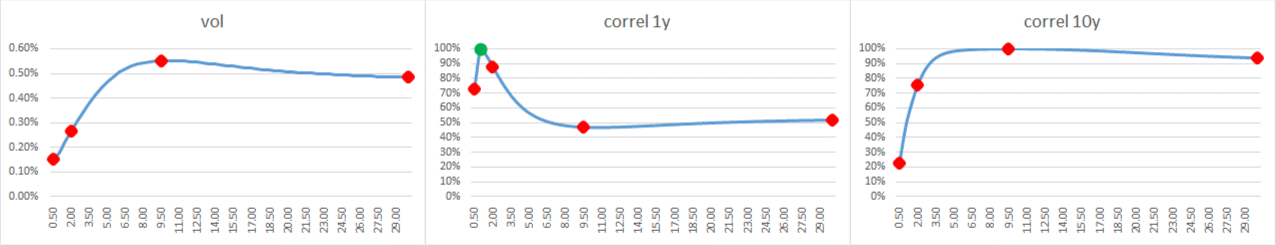
Multi-Factor Cheyette (MFC): calibration

- We fix the n kappas and pick n skew rates with tenors spanning the curve
- We know that the covariance of skew rates in the model is: $\rho_{f_i,f_j}\sigma_{f_i}(t)\sigma_{f_j}(t)=\sum_{l=1}^{n}\sum_{m=1}^{n}\rho_{lm}(t)\lambda_l(t)\lambda_m(t)\exp\left(-k_l\tau_i-k_m\tau_j\right)$
- Inversely: $\begin{bmatrix}\ldots\\ \rho_{lm}\lambda_l\lambda_m\\ \ldots\end{bmatrix}(t)=\begin{bmatrix}\ldots & \ldots & \ldots\\ \ldots & \exp\left(-k_l\tau_i-k_m\tau_j\right) & \ldots\\ \ldots & \ldots & \ldots\end{bmatrix}^{-1}\begin{bmatrix}\ldots\\ \rho_{f_i,f_j}\sigma_{f_i}\sigma_{f_j}\\ \ldots\end{bmatrix}(t)$
- We parameterize the model with the volatilities and correlations of the skew tenors
  And find corresponding factor parameters with the matrix inversion
- Correlation of skew tenors: set to historical
- Volatility of skew tenors:
  - Historical calibration: set to historical flat
  - Implied calibration: set the n lambdas (t) to match n option prices (or best fit more) by expiry

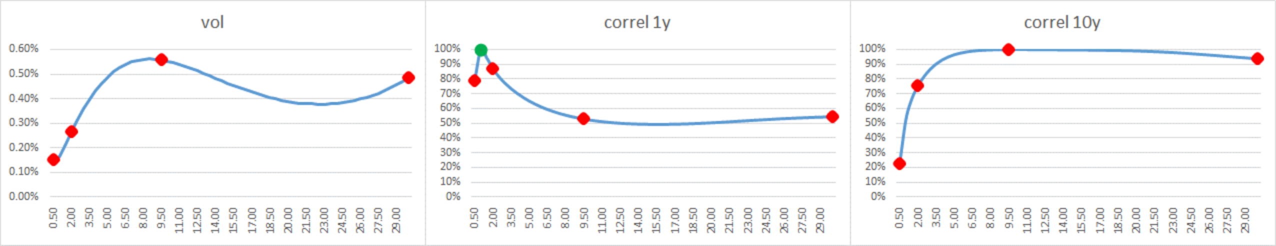
MFC: example

- EUR, 4 factors
- Skew rates: 6m, 2y, 10y, 30y
- Covariance of skew rates fed with 2017 historical estimates
- We compare resulting model with sets of kappas
- We confirm that the impact of kappas vanishes as n grows
- Because kappas act as interpolation coefficients for volatility and correlation structure
- (Probably should consider 5th skew tenor = 20y)

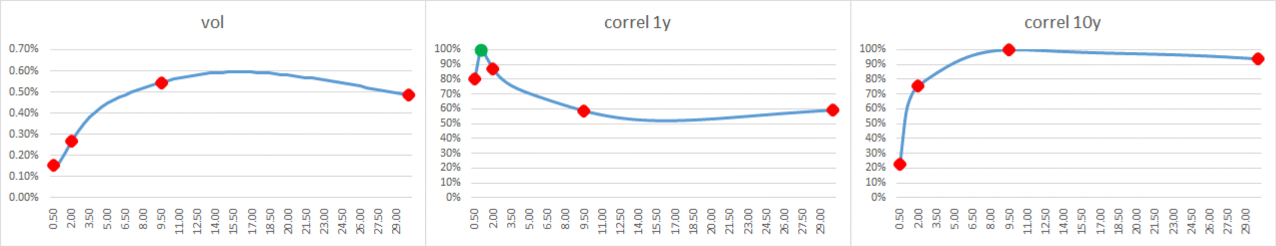
- kappas = (0, 10%, 30%, 100%)



- kappas = (-10%, 10%, 25%, 50%)



- kappas = (5%, 10%, 15%, 20%)



Conclusion

- Interest Rate modelling very active 1990-2005
  - Academic activity culminating with HJM, 1992 and BGM, 1995
  - Practical implementation in financial institutions
    - Multi-factor parametrization and calibration
    - Extensions to local and stochastic volatility to fit market implied volatility smile
    - Numerical implementation
  - Supporting trillion size business 2000-2008
  - From 2005, focus moves to credit and from 2008, regulations
- Two families of IRM emerged
  - Non-Markov family : BGM (1995)
    - Fits historical or implied or mix covariance structure
    - Black box operation
    - Expensive simulation
  - Markov family : Cheyette (1992)
    - Only fits partial historical/implied/mix covariance structure
    - Controlled operation
    - Efficient simulation
- Multi-Factor Cheyette (2005)
  - Combines benefits of BGM and Cheyette families
  - An “efficient BGM” thanks to rate covariance to factor covariance mapping
  - Widely considered the peak interest rate model