The Swaption Cube Prediction Problem

Marc Chataigner - Europlace Institute of Finance
Stéphane Crépey - LaMME, Univ. Evry, CNRS, Université Paris-Saclay

Financial context

- Swaps :
  - Financial derivatives exchanging floating rates against fixed rate during T years (tenor).
  - The Swap rate corresponds to the fixed rate $S(T)$ such that the initial swap price $= 0$.
  - The forward swap rate $F(U,T)$ is the swap rate S but for a swap starting at a future date U.

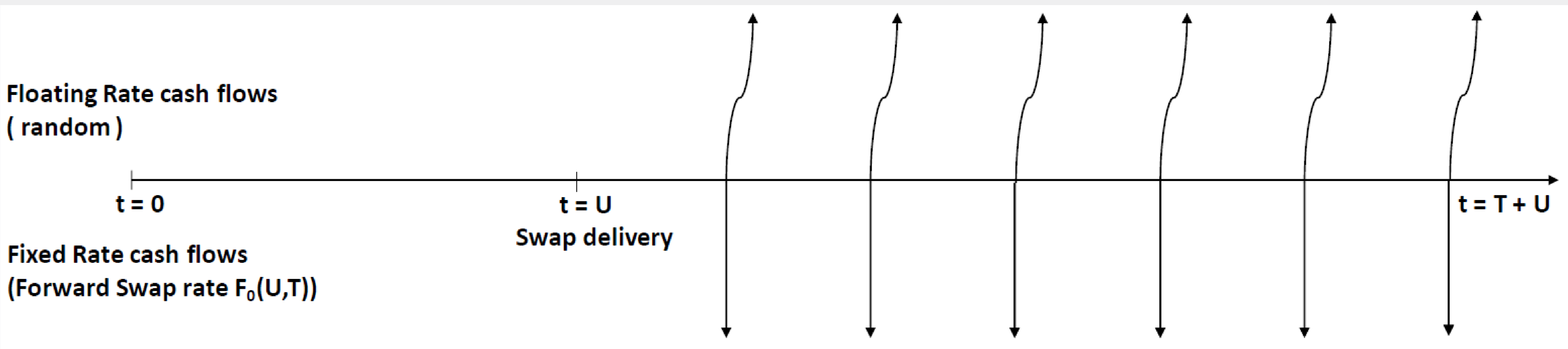

Figure: Payoff scheme for forward swaps

- Swaption :
  - Right to enter into a swap at time U (expiry) with predefined strike K.
  - Can be seen as a call/put option on the underlying forward swap rate.
  - Implied volatility σ matching Black model price with market price.
- Swaption (implied volatility) cube :
  - σ defined along three dimensions (U,T,K).
  - Price convex w.r.t K and increasing w.r.t. $\sigma \rightarrow$ strong no-arbitrage relationships.
  - Cash flows overlapping for close $(U,T) \rightarrow$ statistical arbitrage.

Problem at hand

- HSBC trading desks face data access constraints :
  - Swaption smile : data providers refresh implied volatilities once a day.
  - Forward swap rates : refreshed on an intraday basis.
  - Traders would like to have an intraday view of implied volatilities.
- Connection between several academic fields :
  - Financial mathematics : no-arbitrage relationship, dynamics for implied volatilities, ...
  - Data science : features engineering (missing data, ...), model selection, hyperparameters calibration, ...
  - Computer science : dealing with datasets size, calibration speed, ...

Formalization of the problem

- Input data :
  - Past and current forward swap rate surfaces.
  - Past swaption implied volatility cubes.
- Output data :
  - Current implied volatility cube.

With d_T the number of tenor points, d_U the number of expiry points, d_K the number of strike points, and n_d the number of days in our history :
Our learning problem amounts to finding an application

$$f : \mathbb{R}^{(n_d-1)\times d_K\times d_T\times d_U + n_d\times d_T\times d_U} \ni X \rightarrow Y \in \mathbb{R}^{d_K\times d_T\times d_U}$$

that minimizes the error criterion

$$\sqrt{\frac{1}{d_K\times d_T\times d_U}\sum_{\text{T=h, U=i, and K=j}}\left(Y^{h,i,j} - f(X)^{h,i,j}\right)^2}$$

$f(X)^{h,i,j}$: Swaption implied volatility prediction with cube coordinate h,i,j.

Simplified problem

- Input data:
  - Today's and yesterday's forward swap rate surfaces.
  - Yesterday's swaption implied volatility cube.

$$f : \mathbb{R}^{d_K\times d_T\times d_U + 2d_T\times d_U} \ni X \rightarrow Y \in \mathbb{R}^{d_K\times d_T\times d_U}$$

Modified error criterion :

$$\sqrt{\frac{1}{\text{nbScenarios}}\sum_{\text{scenarios}}\left(\frac{1}{d_K\times d_T\times d_U}\sum_{\text{T=h, U=i, and K=j}}\left(Y^{h,i,j} - f(X)^{h,i,j}\right)^2\right)} \qquad (1)$$

- Simulated data with independent scenarios.

	Input data						Output data	
	Previous smile cube		Previous Forward matrix		Current forward matrix		Current smile cube	
Diffusion model 1	(S_h,T_i,K_j)	...	(S_h,T_i)	...	(S_h,T_i)	...	(S_h,T_i,K_j)	...
Scenario 1								
⋮								
Scenario n								

Figure: Datasets structure.

- 10^4 scenarios generated from a pricing model.
- Model parameters calibrated on one day of real market data.

Three approaches

To cope with the high dimensionality of the learning problem (of f), 3 different approaches have been implemented :

1. *Mapping with forwards* :
   - For each fixed strike K, find a non linear relationship between the forward swap rate surface and the implied volatility surface $(\sigma(U,T,K))$,
     $$f_1 : \mathbb{R}^{d_T\times d_U} \rightarrow \mathbb{R}^{d_T\times d_U}$$
   - f_1 parameterized by multilayer perceptron (MLP) with tanh activation function, 2 hidden layers, and 25 hidden units.
   - Repeating the estimation for each strike $K \rightarrow f \sim d_K$ regressors $\hat{f}_1$.
2. *Nearest neighbors regressions* :
   - Addressing statistical arbitrage w.r.t (U,T);
   - Regularizing the predicted cube with a local regression (for fixed K).
     $$\underset{(\alpha_0,\beta_0)}{\arg\min}\sum_{i=1}^{N} D(\|(U,T)_i-(U,T)_0\|)(\sigma(U,T,K)_i-\alpha_0-\beta_0 F(U,T)_i)^2$$
     $$f_2 : \mathbb{R} \ni F(U,T)_0 \rightarrow \alpha_0 + \beta_0 F(U,T)_0 \in \mathbb{R}$$
     - Number of neighbors N is a function of (U,T) and D is a kernel depending on overlapping cashflows.
   - Repeating the estimation of (α,β) for each cube point $\rightarrow f \sim d_K\times d_T\times d_U$ regressors $\hat{f}_2$.
3. *Skew dynamics* :
   - For each fixed (U,T), today's skew $\sigma(K)$ derived from yesterday's skew, today's and yesterday's underlying forward swap rates:
     $$f_3 : \mathbb{R}^{(d_K+2)} \rightarrow \mathbb{R}^{d_K}$$
   - Also parameterized by MLP (cf Kondratyev, "Learning curve dynamics with artificial neural networks", 2018), with tanh activation function, 2 hidden layer, and 21 hidden units.
   - Repeating the estimation for each pair $(U,T) \rightarrow f \sim d_T\times d_U$ regressors $\hat{f}_3$.

Numerical results

Heatmaps representing the root mean squared absolute error of swaption implied volatilities as a function of the log-moneyness ($\log\frac{F(U,T)}{K}$) and (U,T) :

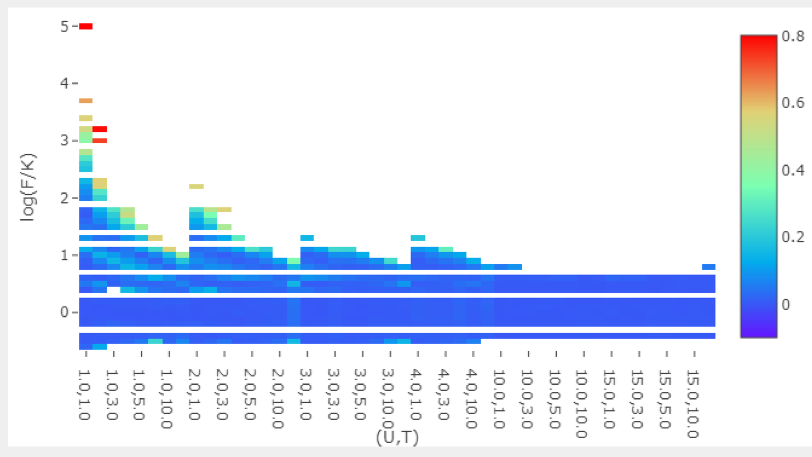

Figure: Heatmap error for *Mapping with forwards*, testing error = 4.9 %

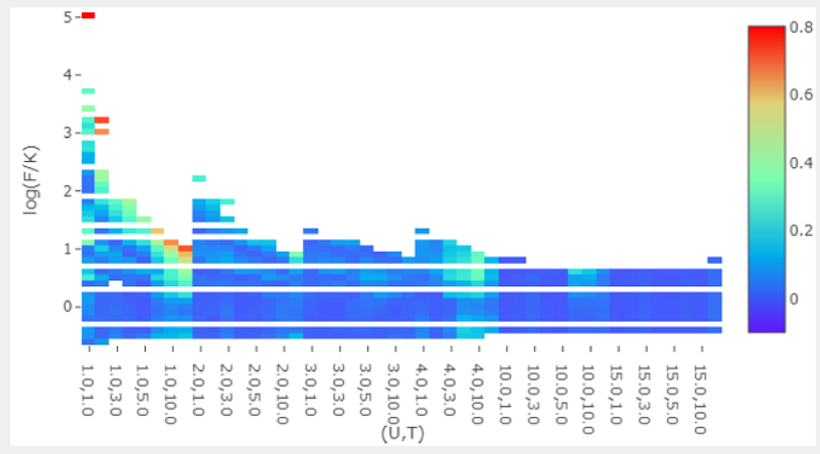

Figure: Heatmap error for *Nearest neighbors regressions*, testing error = 5.8 %

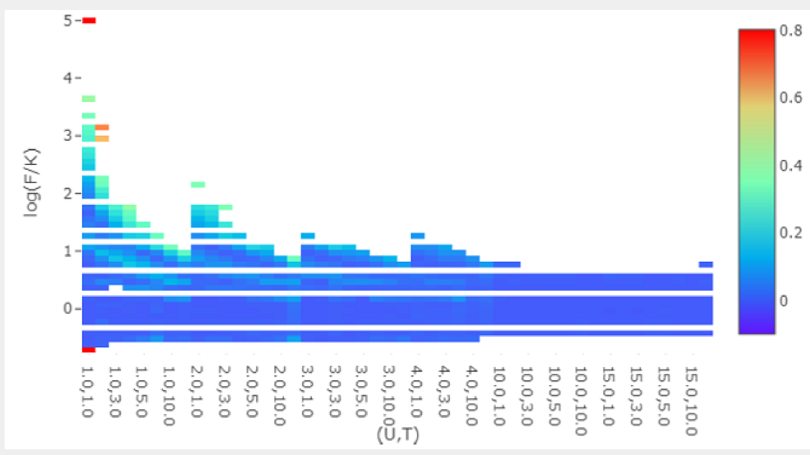

Figure: Heatmap error for *Skew dynamics*, testing error = 0.49%

Error patterns common to the 3 schemes :

- Increasing w.r.t T, decreasing w.r.t U.
- Outliers on the boundary (deep in the money).
- *Skew dynamics* approach more accurate.
- *Nearest neighbors regressions* approach smoother on the boundary.

Absolute testing errors in captions are given by Eq. (1).

Next steps

- Generating a viable implied volatility cube :
  - Statistical arbitrage over (U,T) can be dealt with by regularization or validation procedures.
  - Stratifying estimation with market regimes (hidden markov models, ...).
  - Using autoencoder to capture an invariant structure of implied cube accross time.
- Initial learning problem should be adapted to a time series setup:
  - Backtesting on real market data.
  - Dealing with missing/dirty data.
  - Recurrent Neural Network (LSTM).

This research benefited from the support of the "Chair Modelling of equity and derivative markets" under the aegis of Louis Bachelier laboratory, a joint initiative of HSBC and Collège de France. The research study of M. Chataigner is funded by Europlace Institute of Finance and the Chair Modelling of equity and derivative markets.
Special thanks to the students participants of the 2018 Fields-China joint industrial problem solving workshop in finance, Shanghai Jiao Tong University.