Slides Projet Machine Learning en Finance

Dylan Riboulet

22 Mai 2024

Contexte: Post-Crise vs Pre-Crise

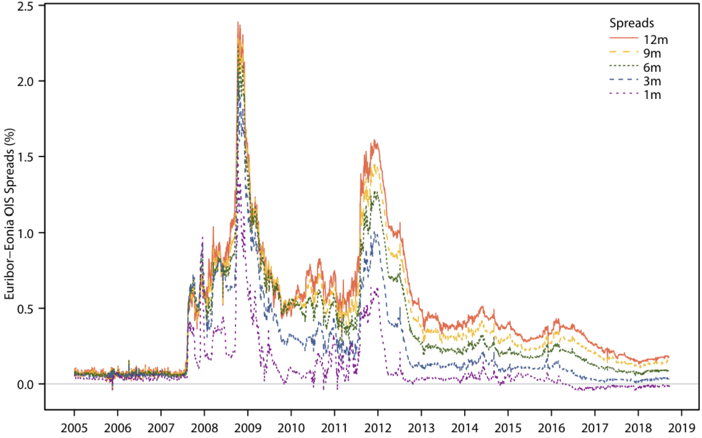

Figure: Euribor - Eonia OIS Spread pour différentes maturités (1 mois à 12 mois) de Janvier 2005 - Septembre 2018. Source: Bloomberg et European Central Bank.

Partie Théorique - Modèles à structure affine

- Considérons l'EDS donnée par:

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t, \quad X_0 > 0,$$

où W est un mouvement brownien 1-dimensionnel sous $\mathbb{Q}$, et

$$\mu(t, r) = \alpha(t) + \beta(t)r, \quad \sigma(t, r) = \gamma(t) + \delta(t)r,$$

avec $\alpha, \beta, \gamma, \delta$ étant des fonctions déterministes, continues sur $\mathbb{R}_+ \times \mathbb{R}$, et l'EDS admet une solution unique.

- Sous ces conditions, nous avons:

$$\mathbb{E}\left[\exp\left(-\int_t^T X_s\, ds - \alpha X_T\right) \mid \mathcal{F}_t\right] = e^{A(t,T) - C(t,T)X_t},$$

pour $0 \leq t \leq T$ et $\alpha \in \mathbb{R}$, où $A(t, T)$ et $C(t, T)$ sont des fonctions déterministes.

Modèle de Vasicek

- EDS pour le processus stochatisque de taux d'intérêt :

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

- r_0 : taux d'intérêt initial
- W : Q-mouvement Brownien, avec Q la mesure de probabilité risque-neutre
- κ : taux de retour à la moyenne
- σ : paramètre de volatilité
- θ : moyenne à long terme
- Distribution de r_t :

$$r_t \sim \mathcal{N}\left(r_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t}), \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t})\right)$$

- Équations de Riccati

$$C(t, T) = \frac{1}{\kappa}(1 - e^{-\kappa(T-t)})$$

Modélisation Multi-Courbe

- r Processus d'Ornstein-Uhlenbeck bi-dimensionnel :

$$dr_{i,t} = \kappa_i(\theta_i - r_{i,t})\, dt + \sigma_i\, dW_{i,t}, \quad t \geq 0, \quad i = 1, 2,$$

où W_1, W_2 sont deux mouvements browniens standards sous Q avec une corrélation ρ.

- Les prix des obligations zéro-coupon :

$$B(t, T) = e^{-A(t,T) - C(t,T) r_{1,t}}$$

- Le taux d'intérêt pour le tenor δ est $r_1 + r_2$. Cela implique que :

$$B(t, T, \delta) = \mathbb{E}^Q \left[e^{-\int_t^T (r_{1,s} + r_{2,s})\, ds} \mid \mathcal{F}_t \right]$$

Modélisation Multi-Courbe (Suite)

- **Fonction Multi-Courbe :**

$$B(t, T, \delta) = \exp\left(\Phi(T - t) + \Psi(T - t)^\top r_t\right), \text{ pour } t \leq T$$

- **Composantes de Ψ :**

$$\Psi_1(T - t) = -\frac{1}{\kappa_1}(1 - e^{-\kappa_1(T-t)}),$$
$$\Psi_2(T - t) = \frac{1}{\kappa_2}(1 - e^{-\kappa_2(T-t)}),$$

- **Fonction Φ**

Régression par Processus Gaussien (GPR)

- **Modèle de Régression :**
  - Chaque observation est modélisée par :

    $$y_i = f(x_i, \theta) + \epsilon_i, \quad i = 1, \ldots, n,$$

    où $f(\cdot, \theta)$ est une fonction de régression dépendant des paramètres inconnus θ, et ϵ est un vecteur de bruit de dimension d.
  - Le bruit ϵ_i consiste en des erreurs gaussiennes i.i.d. avec une moyenne nulle et un écart-type $\hat{\sigma}$.
- **Méthode Bayésienne :**
  - GP Prior
- **Distribution Postérieure à l'aide de la formule de Bayes:**

$$p(\theta \mid x, y) = \frac{p(\theta, y \mid x)}{p(y \mid x)} = \frac{p(y \mid x, \theta)p(\theta)}{p(y \mid x)}, \tag{1}$$

GPR pour la prédiction des log bond prices

- D'après la théorie, en prenant $f = \log$, le processus des *log delta-bond prices* est un processus gaussien. En prenant $x = B(t, T, \delta)$ et $x' = B(s, T, \delta)$, on a
$$f \sim \mathrm{GP}(m, k),$$
  - Sa fonction moyenne :
$$m(x) = \mathbb{E}[f(x)],$$
  - Sa fonction de covariance, appelée noyau (kernel) :
$$k(x, x') = \mathrm{cov}(f(x), f(x')),$$

GPR pour la prédiction des log bond prices (Suite)

- **Fonction Moyenne :**

$$\mu(t,T) := \mathbb{E}[\log B(t,T)] = -A(t,T) - C(t,T)(r_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t}))$$

où A, C, κ, et θ sont des paramètres du modèle.

- **Fonction de Covariance :**

$$c(s,t,T) := \mathbb{E}[\log B(s,T) - \mathbb{E}[\log B(s,T)]] \cdot [\log B(t,T) - \mathbb{E}[\log B(t,T)]]$$

- **Observations avec Bruit :**
  - Les observations consistent en un vecteur $\mathbf{y} = (y(t_1,T),\ldots,y(t_n,T))^\top + \epsilon$ des log bond prices avec un bruit gaussien i.i.d. de variance $\hat{\sigma}^2$.
  - En conséquence, $\mathbf{y} \sim \mathcal{N}(\mu_{\mathbf{y}}, \Sigma_{\mathbf{yy}})$ où :

$$\mu_{\mathbf{y}} = (\mu(t_1,T),\ldots,\mu(t_n,T))^\top$$

$$(\Sigma_{\mathbf{yy}})_{i,j} = c(t_i,t_j,T) + \hat{\sigma}^2 \mathbf{1}_{\{i=j\}}$$

Distribution Conditionnelle de ξ sachant y

- Considérons deux vecteurs aléatoires ξ et y suivant une distribution normale multivariée.
- La distribution conditionnelle de ξ sachant y est donnée par la formule :

$$\xi \mid y \sim \mathcal{N}\left(\mu_{\xi|y}, \Sigma_{\xi|y}\right),$$

où :

$$\begin{aligned}\mu_{\xi|y} &= \mu_\xi + \Sigma_{\xi y}\Sigma_{yy}^{-1}(y - \mu_y), \\ \Sigma_{\xi|y} &= \Sigma_{\xi\xi} - \Sigma_{\xi y}\Sigma_{yy}^{-1}\Sigma_{y\xi}.\end{aligned}$$

Prédiction et Distribution Conditionnelle

- **Données d'Entraînement et Prédiction :**
  - Soit les données d'entraînement $\mathbf{y} = (y(t_1, T), \ldots, y(t_n, T))^\top$.
  - Objectif : prédire le vecteur $\tilde{\mathbf{y}} = (y(s_1, T), \ldots, y(s_m, T))^\top$.
- **Distribution Jointe :**

$$\begin{bmatrix} \mathbf{y} \\ \tilde{\mathbf{y}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_{\tilde{y}} \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & \Sigma_{y\tilde{y}} \\ \Sigma_{\tilde{y}y} & \Sigma_{\tilde{y}\tilde{y}} \end{bmatrix} \right),$$

  où $\boldsymbol{\mu}_y$ et Σ_{yy} sont les vecteurs moyenne et noyau des observations
- **Détails des Composantes :**

$$(\boldsymbol{\mu}_{\tilde{y}})_i = \mu(s_i, T), \quad (\Sigma_{y\tilde{y}})_{i,j} = c(t_i, s_j, T)$$

$$(\Sigma_{\tilde{y}y})_{i,j} = c(s_i, t_j, T), \quad (\Sigma_{\tilde{y}\tilde{y}})_{i,j} = c(s_i, s_j, T)$$

- **Distribution Conditionnelle :**

$$\tilde{\mathbf{y}} \mid \mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_{\tilde{y}|y}, \Sigma_{\tilde{y}|y})$$

  Obtenu via l'équation précédente, conditionnellement aux données d'entraînement $\mathbf{y}$.

Distribution Postérieure

- **Prédicteur moyen :** $\hat{m}(x_*)$ pour $f(x_*)$:

$$\hat{m}(x_*) := \mathbb{E}[f(x_*) \mid y] = \mathbb{E}[y_* \mid y] = \mu_{\tilde{y}|y} = \mu_{\tilde{y}} + \Sigma_{\tilde{y}y}\Sigma_{yy}^{-1}(y - \mu_y),$$

  et estimation de la variance pour l'intervalle de confiance :

$$\mathrm{Var}(y_* \mid y) = \Sigma_{\tilde{y}|y} = \Sigma_{\tilde{y}\tilde{y}} - \Sigma_{\tilde{y}y}\Sigma_{yy}^{-1}\Sigma_{y\tilde{y}}$$

- **Application des Données d'Entraînement**

Illustration des Prédictions avec distribution Prior

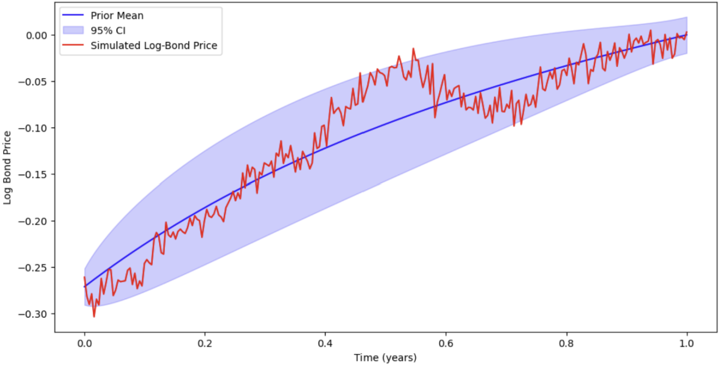

Figure: Illustration de la distribution prior des log bond prices, conditionnées sur 176 exemples d'entraînement simulés.

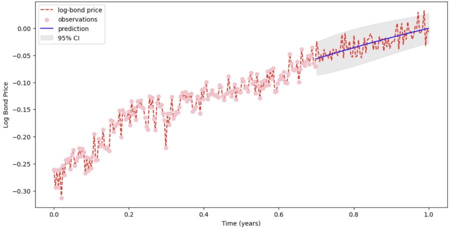

Figure: Illustration de la distribution postérieur conditionnées sur 70% d'observations passées.

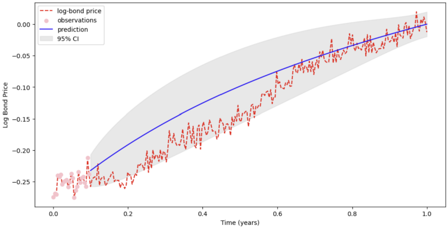

Figure: Illustration de la distribution postérieur conditionnées sur 10% d'observations passées.

Évaluation des prédictions

- **Mean Squared Error (MSE):**

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \tilde{y}_i)^2 \tag{2}$$

- **Standardized Mean Squared Error (SMSE):**

$$\text{SMSE} = \frac{1}{\tilde{\sigma}^2}\frac{1}{m}\sum_{i=1}^{m}(\tilde{y}_i - \tilde{\mu}_i)^2 \tag{3}$$

- **Résultats Obtenues**

```
Mean Squared Error (MSE): 0.00032736984056532746
Standardized Mean Squared Error (SMSE): 0.7492729170203547
```

Prédiction des Log-Delta-Bond-Prices

- **Objectif:** Prédire les log-bond prices non observés avec 176 données d'entrainements $T = 252$ days.

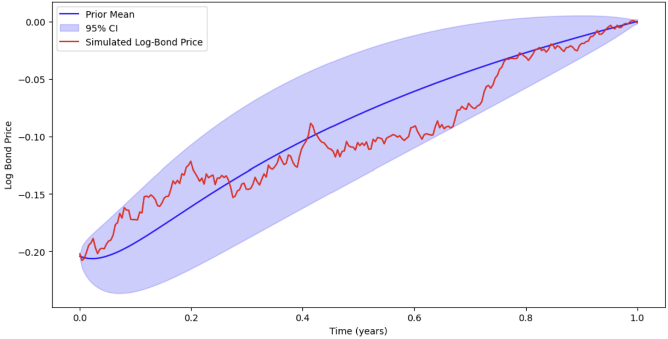

Figure: Illustration du prior du GP. La ligne bleue représente la moyenne du prior, alors que la zone ombrée est l'IC à 95%

Cadre Multi-Courbe pour la prédiction et calibration

- **Distribution Normale des Observations :**
  - Observations de deux prix d'obligation pour chaque point temporel t_i : $B(t_i, T, 0)$ et $B(t_i, T, \delta)$.
  - Le vecteur des observations $\mathbf{y} = (\mathbf{y}_0, \mathbf{y}_\delta)^\top$ est distribué normalement, $\mathbf{y} \sim \mathcal{N}(\mu_{\mathbf{y}}, \Sigma_{\mathbf{y}})$, avec :

$$\mu_{\mathbf{y}} = \begin{bmatrix} \mu(\cdot, T, 0) \\ \mu(\cdot, T, \delta) \end{bmatrix}, \Sigma_{\mathbf{y}} = \begin{bmatrix} \Sigma^{\mathbf{y}}_{00} & \Sigma^{\mathbf{y}}_{0\delta} \\ \Sigma^{\mathbf{y}}_{\delta 0} & \Sigma^{\mathbf{y}}_{\delta\delta} \end{bmatrix}$$

- **Méthodologie de Calibration**

Simulation des bond prices et delta-bond prices

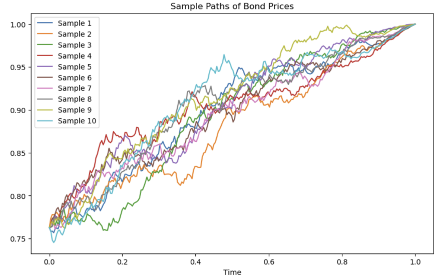

Figure: 10 des 1000 échantillons de prix d'obligations zero coupons simulés

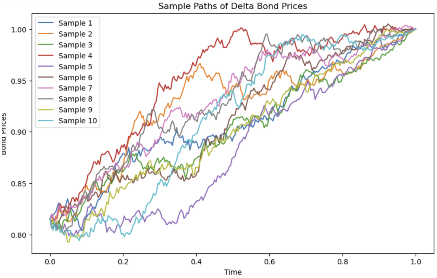

Figure: 10 des 1000 échantillons de prix d'obligations delta zero coupons simulés

Sélection des Hyperparamètres

- **Fixation des Paramètres et Valeurs Réelles :**
    - Initialisation aléatoire et $r_0 = 0.5$, $\kappa = 2$, $\theta = 0.1$, $\sigma = 0.2$.
- **Optimisation sur des Échantillons :**
    - Pour chacun des 1 000 échantillons, nous cherchons le choix optimal des hyperparamètres $\Theta = \{r_0, \kappa, \theta, \sigma\}$ basé sur la trajectoire simulée de 252 prix des obligations log-bond.
- **Minimisation de la Log-Vraisemblance Marginale Négative :**

$$\log p(y) = c - \frac{1}{2}(y - \mu_y)^\top \Sigma_{yy}^{-1}(y - \mu_y),$$

Illustration des resultats de la calibration single-curve

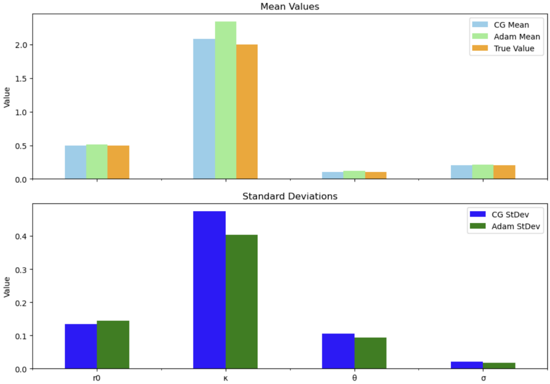

Figure: Hyperparamètres optimaux

Observations : Résultats tres satisfaisant pour les deux algorithmes d'optimisation CG et Adam.

Comparaison des algorithmes d'optimisation CG et Adam pour la calibration multi-courbes

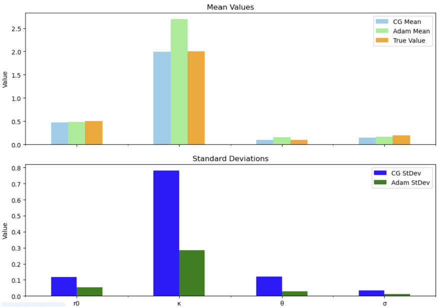

Figure: Hyperparamètres θ_1, κ_1, σ_1, r_{01}

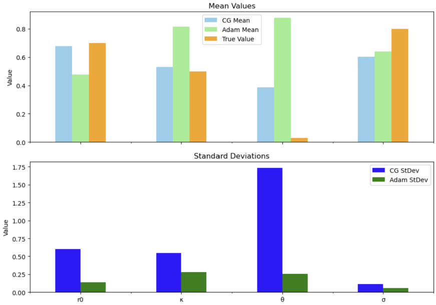

Figure: Hyperparamètres θ_2, κ_2, σ_2, r_{02}

Observations : L'optimiseur CG présente une plus grande variabilité dans les paramètres estimés, particulièrement pour θ_2 dans les configurations multi-courbes, comparé à Adam.

Conclusion des simulations de calibration

- **Comparaison des cadres de calibration :**
  - La calibration dans le cadre multi-courbes est une tâche plus **complexe** que dans le cadre à courbe unique.
- **Considérations computationnelles :**
  - Pour gérer la complexité computationnelle, nous avons utilisé **126 points de données d'entraînement**.
- **Implications pour la performance du modèle :**
  - Augmenter le nombre de points de données d'entraînement pourrait **améliorer la précision du modèle**, mais au prix d'un temps de calcul **accru**.

Conclusion et Perspectives

- **Distribution a posteriori :**
  - Beaucoup plus d'informations par rapport à une prédiction simple
- **Gestion de l'interpolation et du lissage par les GP**
  Entrées d'apprentissage exactes ou bruitées
- **Quantification de l'incertitude :**
  - IC permettant une évaluation précise des risques
- **Interaction avec la formation dynamique :**
  - Utile dans le cas où les modèles doivent s'adapter rapidement à de nouvelles informations

Perspectives de Recherche

- **Analyse Comparative des Estimateurs :**
  - Avec les estimateurs du maximum de vraisemblance par exemple
- **Étude des Techniques d'Optimisation Avancées :**
  - AdaGrad, Adadelta, RMSProp, NAG.
- **Prédiction des Prix des Obligations Zéro Coupon :**
  - Un support décisionnel pour l'achat d'options sur des obligations ZC
  - Méthodes de réechantillonage dans le cas de manque de données
- **Développement à d'autres Modèles de Taux Court :**
  - Cox-Ingersoll-Ross, Hull-White étendu à Vasicek
- **Calibration des Modèles de Taux d'Intérêt avec Sauts :**
  - Ajout d'un Processus de Poisson
- **Etude de la Calibration des Marchés de Taux d'Intérêt :**
  - Réseaux de Neurones Bayésiens

References

1. Sandrine Gumbel and Thorsten Schmidt: *Machine learning for multiple yield curve markets: fast calibration in the Gaussian affine framework*, April 20, 2020.
2. Grbac, Z., Gumbel, S., Fontana, C. and Schmidt, T. (2020), *Term structure modelling for multiple curves with stochastic discontinuities*, Finance Stochastics (online).
3. Cours Master-MLF2324-lect1-HPham.pdf: *Machine learning in finance: Theoretical foundations*, Huyên PHAM