University College London

Department of Computer Science

Macroeconomic Drivers of Implied Volatility in USD and EUR Swaption Markets

Author
Candidate Number: PDWQ3

Academic Supervisor
Professor Guido Germano
Department of Computer Science
University College London

Industrial Supervisor
Mr Kirill Egorov
SilverTide

September 8, 2025

Abstract

Option-implied interest rate volatilities surged after 2021 to their highest levels in over a decade and remain historically elevated. This leads to heightened uncertainty around inflation, growth, and monetary policy, and has direct consequences for investors and traders who rely on volatility measures for pricing and hedging. In this context, the thesis asks whether swaption-implied volatility in USD and EUR can be explained by rate levels alone, or whether broader macroeconomic conditions also play a role.

We study this question using a two-stage framework on daily data from 2014–2024. We first fit a volatility–rate backbone to capture the structural link between yields and volatility, then explain the remaining variation with macro drivers such as forward rates, inflation expectations, and balance sheet policy. To test the strength of these results, we estimate models with both linear and nonlinear methods.

The analysis points to three main results. First, explanatory power depends on the regime. Models fit only moderately across 2014–2024, but perform much better in 2021–2024, when both rates and volatility were high. Second, only a small number of macro drivers matter consistently, with forward rates and balance sheet policy the most important. Third, more complex methods do not improve predictive accuracy, but they help confirm which factors are most relevant across maturities.

Overall, the thesis shows that volatility is not explained by rate levels alone but is also driven by regime-specific macroeconomic factors. Academically, this highlights the need to account for regime and macro conditions in volatility modeling. For practitioners, the framework offers a way to interpret and anticipate volatility patterns across maturities and regimes, with direct applications in option pricing, portfolio risk management, and volatility forecasting.

All code and data preprocessing scripts used in this thesis are openly available at:
`https://github.com/chiarab24/implied-volatility-thesis.git`.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Interest rate volatility is central to fixed income markets. It affects how derivatives are priced, how monetary policy passes through the yield curve, and how risk is managed. While it is well established that volatility is not constant but varies with the level of interest rates, the full set of drivers is less well understood. Prior work documents a ‘universal relationship’ between rates and volatility, with a recurring three-regime pattern observed across currencies, maturities, and time periods [1, 2]. This provides a consistent empirical benchmark for modeling volatility.

Since 2021, option-implied volatilities of interest rates have surged to their highest levels in more than a decade and remain elevated [3]. This reflects renewed uncertainty around inflation, growth, and monetary policy. Such shifts show that recent volatility cannot be fully explained by changes in rate levels, suggesting the influence of other factors. Option-implied volatility also correlates with macroeconomic conditions. For example, Federal Reserve research finds that spikes in swaption-implied volatility tend to precede weaker activity and greater downside risks to GDP growth [4]. Periods of unusually high or low implied volatility are often linked to macro factors such as inflation expectations, central bank balance sheet policy, and liquidity. These influences are not captured by level-dependent models, yet they are crucial for understanding volatility dynamics.

To address this, recent industry work, particularly JP Morgan’s empirical framework, uses a two-step approach: first fitting a volatility–rate relationship, then explaining the remaining variation with macroeconomic factors [5, 6]. This thesis builds on that approach and extends it. Using USD and EUR swaption data from 2014–2024, it asks whether volatility can be explained by rate levels alone or whether macroeconomic forces provide additional explanatory power. The objectives are: (i) to test the robustness of the volatility–rate backbone across currencies and regimes, (ii) to identify which macro drivers matter most for residual volatility, and (iii) to compare linear and nonlinear methods. The results show that volatility is not only dependent on rates but also driven by regime-specific macroeconomic factors that differ between the USD and EUR markets.

1.2 Research objective

This thesis tests whether interest rate volatility can be explained by rate levels alone or whether macroeconomic conditions add systematic explanatory power. To address this, the analysis applies a two-stage framework to USD and EUR swaption data from 2014–2024. In the first stage, a volatility–rate backbone is fitted to describe the empirical relationship between volatility and interest rate levels. This serves as a structural benchmark. In the second stage, the residual variation is modeled using macroeconomic drivers such as forward rates, inflation expectations, central bank balance sheets, swap spreads, and the short-end curve slope.

Recent industry research underlines how relevant this question has become. JPMorgan proposed an empirical framework for USD swaptions that combines a backbone fit with regressions on macro drivers such as policy expectations, liquidity, and balance sheet size [5, 6]. BNP Paribas applied a similar approach to EUR swaptions, linking deviations from the backbone to long-dated yields and inflation swaps [7]. Both demonstrate the influence of macroeconomic forces, but they focus on single currencies and rely on linear methods only.

This thesis extends that work in three ways. First, it identifies which macro drivers explain deviations from the backbone across maturities and regimes in both USD and EUR. Second, it quantifies how much explanatory power macro factors add beyond the backbone, and whether their influence differs by currency. Third, it tests whether regularized linear methods and nonlinear models, such as random forests, offer potential advantages over ordinary least squares. By combining multiple modeling approaches and comparing the two largest swaption markets, the thesis examines how macroeconomic conditions drive interest rate volatility and tests whether more flexible methods improve explanatory power over standard industry models.

1.3 Structure of the thesis

The remainder of the thesis is organized as follows. Chapter 2 reviews the literature on interest rate volatility, with a focus on swaption markets and the empirical link between volatility and rate levels. It also introduces the machine learning methods and evaluation metrics used later in the analysis. Chapter 3 describes the data set, the two-stage framework, and the set of macroeconomic drivers included in the regressions. Chapter 4 presents the empirical results, covering OLS baselines, feature selection methods, and comparisons across currencies and time windows. Chapter 5 interprets these results in light of macroeconomic intuition, highlighting regime dependence and differences between USD and EUR. Chapter 6 concludes with a summary of the main findings, while Chapter 7 outlines directions for future research.

2 Literature Review

This chapter sets the background for the thesis. It begins with interest rate volatility: how it is defined, how swaptions quote it, and how it tends to move with the level of rates. The discussion then turns to modeling approaches, from constant-volatility frameworks to local and stochastic volatility models, before arriving at the SABR model. The second part introduces core concepts from machine learning, such as the bias–variance trade-off, linear and nonlinear regression, and feature selection. The chapter ends with evaluation metrics, including mean squared error, out-of-sample R^2, and Pearson correlation. Together these elements provide the foundation for the two-stage volatility framework developed in the thesis.

2.1 Volatility in interest rate markets

Volatility is a fundamental concept in interest rate derivatives, as it captures the uncertainty surrounding future rate movements. It enters directly into the valuation of options, determines hedging ratios, and provides a measure of risk for trading and portfolio management [8]. Unlike in equity markets, where volatility is often treated as a single number, fixed income markets rely on a more nuanced set of conventions reflecting the specific features of interest rates, such as mean reversion, the possibility of negative levels, and the dependence of volatility on the rate environment [8, 9].

In practice, volatility in rates can be approached from two perspectives. Historical volatility is based on the realized variability of past rate changes, while implied volatility is extracted from option prices and reflects market expectations of future uncertainty. For swaptions and other rate options, implied volatility has become the dominant measure, as it allows market participants to compare option values across maturities and strikes without directly referencing option prices [9].

The following sections explain how implied volatility is extracted, outline the pricing models and quoting conventions in swaption markets, and review empirical approaches to modeling interest rate volatility.

2.1.1 Implied volatility and its extraction

In fixed income markets, volatility is described in two main ways: historical volatility and implied volatility. Historical volatility is a backward-looking measure, which is computed as the standard deviation of past changes in the rate over a chosen window. It provides a simple summary of realized variability. By contrast, implied volatility is a forward-looking measure inferred from option prices. It is defined as the volatility parameter that, when inserted into an option pricing model, equates the model price with the observed market premium. Since most models are not analytically invertible in volatility, numerical schemes are required to compute implied volatility. In class we saw, and the literature confirms, that the Newton–Raphson method and the approach proposed by Jäckel are two of the most widely used procedures for this task [10, 11, 12]. These methods ensure fast and stable convergence when inverting option pricing formulas, which makes them essential in the construction of volatility surfaces [13].

In the case of swaptions, implied volatility is quoted directly rather than price. Market practice distinguishes between two conventions: lognormal volatility (Black, 1976), where changes are proportional to the rate level, and normal volatility (Bachelier), where changes are absolute [9, 14]. The lognormal volatility dominates in positive-rate environments, while the normal convention is preferred when rates are close to or below zero [9]. This choice affects the scaling and interpretation of volatility and whether option values behave consistently across different rate regimes.

2.1.2 Pricing models and market conventions

Swaptions are options on forward-starting interest rate swaps and are quoted directly in terms of implied volatility rather than option premium [9, 15]. The market convention specifies swaptions in expiry-tenor format, for example a 1yx10y swaption grants the right to enter a ten-year swap starting in one year. This quoting convention allows market participants to compare options across maturities and strikes. Swaption pricing frameworks range from the classical Black (1976) and Bachelier models to modern approaches such as the LIBOR Market Model [16].

The choice of pricing model determines whether volatility is quoted in lognormal (yield) terms or normal (basis point) terms. The Black–Scholes model (1973) [17] describes asset prices as a geometric Brownian motion (GBM), implying lognormal distributions. Building on this, the Black (1976) model adapted the framework to forwards and futures, and is the market standard for swaptions in positive-rate environments [9]. In this case, the forward swap rate F_t follows a geometric Brownian motion (GBM),

$$\frac{dF_t}{F_t} = \sigma_{\text{GBM}}\, dW_t, \tag{2.1}$$

where σ_{GBM} is the volatility parameter of the GBM and W_t is a standard Brownian motion. The value V_{Black} of a European payer swaption is given by

$$V_{\mathrm{Black}} = F_0 F_X(d_1) - K e^{-rT} F_X(d_2), \qquad d_{1,2} = \frac{\log(F_0/K) \pm \frac{1}{2}\sigma_{\mathrm{GBM}}^2 T}{\sigma_{\mathrm{GBM}}\sqrt{T}}, \tag{2.2}$$

where F_0 is the forward rate, K the strike price, T the time to expiry, and $F_X(\cdot)$ is the cumulative distribution function (CDF) of the standard normal random variable X [14].

When rates are close to or below zero, the lognormal assumption becomes problematic, as it rules out negative values. In such environments, swaptions are quoted and risk-managed using the Bachelier model. In this case, the forward rate F_t follows an arithmetic Brownian motion (ABM),

$$dF_t = \sigma_{\mathrm{ABM}} dW_t, \tag{2.3}$$

where σ_{ABM} is the volatility parameter of the ABM and W_t is a standard Brownian motion. The value $V_{\mathrm{Bachelier}}$ of a European payer swaption is

$$V_{\mathrm{Bachelier}}(K) = (F_0 - K) F_N(d) + \sigma_{\mathrm{ABM}}\sqrt{T}\, f_N(d), \qquad d = \frac{F_0 - K}{\sigma_{\mathrm{ABM}}\sqrt{T}}. \tag{2.4}$$

where F_0, K, and T have the same meaning as above, $F_N(\cdot)$ is the CDF of the standard normal distribution N, and $f_N(\cdot)$ its PDF [9].

The coexistence of these two conventions shows how rate dynamics change across regimes. In the lognormal approach, rate moves are proportional to the level of rates, so higher rates imply larger moves. In the normal approach, rate moves are treated as absolute changes, which works better when rates are close to zero or negative. In practice, the market switches between the two so that swaption prices remain meaningful and consistent across different rate environments [9].

A final step is the extraction of implied volatility from observed swaption premiums, known as volatility inversion. Because neither the Black nor the Bachelier formula can be inverted in closed form, numerical root-finding methods are required. Both the Newton-Raphson method and Jäckel's rational approximation are widely used in practice, and were also covered in class [10, 11, 13]. For the Bachelier model, the inversion problem is computationally simpler than in the Black case, and several efficient approximation methods have been proposed in the literature [9, 18]. These approaches are widely applied in practice for large-scale construction of swaption volatility surfaces.

2.1.3 Volatility surfaces and quoting conventions

For swaptions, implied volatilities are quoted across option expiries and swap tenors, forming what is known as the swaption volatility cube. Each point gives the implied volatility for a specific expiry–tenor pair, such as a 3mx10y swaption. Fixing the strike, which is typically done at-the-money, reduces the cube to two dimensions, expiry and tenor, commonly shown as the volatility surface. In contrast, in equity options, volatilities are typically organized directly as a surface (expiry $\times$ strike) [19], since there is no tenor dimension.

In practice, the Black (1976) and Bachelier models are primarily used as reporting models to express option prices in terms of their implied volatilities. This convention allows market participants to compare the option values across maturities and strikes without directly referencing premiums [9]. Quoting in volatility terms has therefore become the standard language of the market.

Empirically, volatility surfaces show clear patterns. Variation with strike gives the volatility smile or skew, while variation with maturity defines the term structure. These features reflect market views on tail risk, mean reversion, and regime shifts, and contrast with the constant-volatility assumption of the Black–Scholes framework [19].

In this thesis, the volatility surface serves as the reference against which rate-dependent patterns and macroeconomic drivers are studied. Looking across different expiry–tenor combinations allows us to compare how short- and long-dated swaptions evolve across time and regimes.

2.1.4 Empirical structure of interest rate volatility

Equity prices are often modeled as geometric Brownian motions with constant volatility, whereas interest rates are usually treated as mean-reverting processes whose volatility depends on the rate level [19]. Empirical studies show that this dependence is nonlinear, with a pattern that is consistent across currencies, maturities, and time periods. Deguillaume, Rebonato, and Pogudin [1] document a universal three-regime structure. When rates are very low, below two percent, volatility scales proportionally with the rate. In the intermediate range, between roughly two and six percent, volatility is largely independent of the level. Once rates exceed six percent, volatility again increases with the rate. This nonlinear relationship, which holds across different markets and tenors, is illustrated in Figure 2.1.

Rebonato and El Aouadi [2] extended these results, showing that the same piecewise behavior is also present in negative-rate environments. After rescaling rate moves to remove seasonal and time effects, they found that the remaining variability still scaled systematically with the level of the rate, which reinforced the idea of a universal relationship.

This universal rate-volatility relationship has important implications for theory and practice. It challenges models that rely on constant volatility and instead supports approaches that build rate dependence directly into the dynamics. It also follows industry practice, where a structural vol–rate curve is fitted as the first step in empirical models

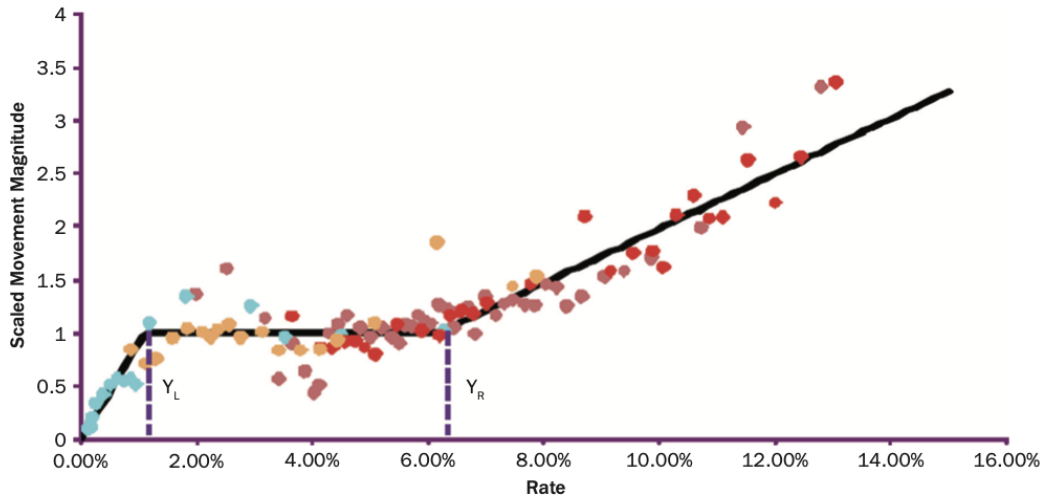

Figure 2.1: Rescaled magnitude of rate moves as a function of the rate level, illustrating the universal three-regime volatility pattern [2, Exhibit 1]. Volatility scales proportionally with the rate when levels are very low (below 2%), becomes independent of the rate in the intermediate range (2-6%), and increases with the rate again once levels exceed about 6%.

[5, 6]. In this thesis, this motivates the first stage of our framework, where we first estimate the volatility–rate backbone empirically, before analyzing the role of macroeconomic factors.

2.1.5 Volatility modeling approaches

The strong link between interest rate levels and volatility has motivated a variety of modeling approaches. These range from the original constant volatility models to more flexible local and stochastic specifications, as well as hybrid frameworks such as SABR. In what follows, we first present models in the standard equity-style notation with S_t, and then show how they extend naturally to the forward rate F_t relevant in interest rate derivatives.

Constant volatility

The Black-Scholes model [17] assumes that volatility is constant and independent of both time and the underlying asset. The asset price S_t follows a geometric Brownian motion,

$$dS_t = (r - q)S_t\,dt + \sigma S_t\,dW_t, \tag{2.5}$$

where r is the risk-free rate, q the continuous dividend yield, and σ the constant volatility. While this assumption yields closed-form option pricing formulas, it is inconsistent with volatility smiles and term structures [20]. This makes the model unsuitable for interest rate derivatives, where volatility clearly depends on rate levels.

Local volatility

Local volatility models, introduced by Dupire [21], relax the constant assumption by letting volatility vary deterministically with both time and state. For an asset price S_t, the stochastic differential equation becomes

$$dS_t = (r - q)S_t\, dt + \sigma_{\mathrm{loc}}(t, S_t)\, S_t\, dW_t^{(1)}. \tag{2.6}$$

For interest rate forwards, working under the forward measure makes the forward rate a martingale, so the drift term disappears and the stochastic differential equation reduces to

$$dF_t = \sigma_{\mathrm{loc}}(t, F_t)\, dW_t^{(1)}. \tag{2.7}$$

Here σ_{loc} is the local volatility surface calibrated from option prices. This framework ensures an exact fit to option prices and arbitrage-free consistency across strikes and maturities. In the case of interest rates, it aligns with the empirical level-dependent patterns reported in [1, 2]. The limitation is that volatility becomes fully determined by (t, F_t), which does not allow for randomness in volatility [19].

Stochastic volatility

Stochastic volatility models introduce randomness by treating volatility itself as a latent process. A prominent example is the Heston model [22], where the asset price under the risk-neutral measure follows

$$dS_t = (r - q)\, S_t\, dt + \sqrt{v_t}\, S_t\, dW_t^{(1)}, \tag{2.8}$$

$$dv_t = \kappa(\theta - v_t)\, dt + \xi\sqrt{v_t}\, dW_t^{(2)}, \tag{2.9}$$

$$\mathrm{Corr}\big(dW_t^{(1)}, dW_t^{(2)}\big) = \rho, \tag{2.10}$$

with v_t the variance process, $\kappa > 0$ the speed of mean reversion, $\theta > 0$ the long-run variance, $\xi > 0$ the volatility of variance, and $\rho \in (-1, 1)$ the instantaneous correlation between the Brownian increments $dW_t^{(1)}$ and $dW_t^{(2)}$. The mean-reversion term $\kappa(\theta - v_t)$ stabilizes v_t by pulling it back toward θ, preventing unbounded drift.

In the interest rate setting, the same specification is often applied directly to forward rates. For forwards under the forward measure, the rate is a martingale. The drift term therefore vanishes and the dynamics reduces to

$$dF_t = \sqrt{v_t}\, F_t\, dW_t^{(1)}, \tag{2.11}$$

$$dv_t = \kappa(\theta - v_t)\, dt + \xi\sqrt{v_t}\, dW_t^{(2)}, \tag{2.12}$$

$$\mathrm{Corr}\big(dW_t^{(1)}, dW_t^{(2)}\big) = \rho, \tag{2.13}$$

where ρ retains the same interpretation as the instantaneous correlation between the Brownian increments driving the forward rate and the variance process.

These models capture volatility clustering, skewness, and heavy tails, but they are more challenging to calibrate and simulate than local volatility models. They also form the basis for the SABR model, which can be viewed as a stochastic volatility specification for F_t that replaces mean reversion with a lognormal volatility process.

The SABR model

The SABR model (stochastic alpha, beta, rho) of Hagan et al. [15] can be interpreted as a special case of stochastic volatility tailored to forward rates. Its dynamics are

$$dF_t = \sigma_t \, F_t^{\beta} \, dW_t^{(1)}, \tag{2.14}$$

$$d\sigma_t = \nu \, \sigma_t \, dW_t^{(2)}, \tag{2.15}$$

$$\mathrm{Corr}\big(dW_t^{(1)}, dW_t^{(2)}\big) = \rho, \tag{2.16}$$

where $\beta \in [0, 1]$ governs how volatility scales with the forward rate: $\beta = 1$ corresponds to lognormal dynamics, $\beta = 0$ to normal dynamics, and intermediate values interpolate between the two. The parameter $\nu > 0$ controls the volatility of volatility, while $\rho \in (-1, 1)$ is the instantaneous correlation between the Brownian increments driving the forward rate and its volatility.

A key difference from Heston-type models is that SABR does not include mean reversion in the volatility process. This means σ_t can drift to very high levels over long horizons, sometimes called "explosive" behavior. Even with this drawback, SABR is widely used in practice: β introduces level dependence, ρ generates skew, and semi-analytic formulas make calibration efficient. Low values of β also make the model work in negative-rate environments [9]. In this way, SABR is a practical stochastic volatility model for forwards, and its mix of flexibility and speed explains why it remains the market standard.

In summary, local volatility fits option prices exactly but does not allow for randomness in volatility, stochastic volatility introduces more realistic dynamics at the cost of harder calibration, and SABR offers a middle ground. For this reason, SABR remains the standard model used in practice.

2.2 Machine learning models

As part of the second stage of our two-stage framework, we build empirical models to explain the deviations of implied volatility from the fitted backbone. The objective is to capture the impact of macroeconomic and market factors on volatility. While linear regression has been widely applied in the literature, this thesis extends the approach by evaluating whether nonlinear models offer additional explanatory power. For background on these methods, see [23].

2.2.1 Problem setup

We consider a supervised learning setting with data

$$\mathcal{D} = \{(\mathbf{x}_n, y_n)\}_{n=1}^{N}, \tag{2.17}$$

where each feature vector

$$\mathbf{x}_n = (x_{n,1}, x_{n,2}, \ldots, x_{n,p}) \in \mathbb{R}^p \tag{2.18}$$

collects the independent variables, and $y_n \in \mathbb{R}$ is the dependent variable. In machine learning terminology, $\mathbf{x}_n$ are the features and y_n the labels. In the context of this thesis, $p = 5$, as each component corresponds to one of the macroeconomic drivers introduced earlier, being the forward rate, long-term inflation expectation, central bank balance sheet, the 10-year swap spread, and the short-end curve slope. The dependent variable y_n is the observed implied volatility for a given swaption structure on day n.

We assume that the observations are generated according to

$$y_n = f(\mathbf{x}_n) + \epsilon_n \quad \text{with} \quad \epsilon_n \sim \mathcal{N}(0, \sigma^2), \tag{2.19}$$

where $f : \mathbb{R}^p \to \mathbb{R}$ is an unknown function and ϵ_n is an independent noise term.

A statistical or machine learning model provides an approximation $\hat{f}(\mathbf{x}; \boldsymbol{\theta})$ to the true function f, parameterized by $\boldsymbol{\theta}$. The model's predictions are then given by

$$\hat{y}_n = \hat{f}(\mathbf{x}_n; \boldsymbol{\theta}), \tag{2.20}$$

and the residuals are defined as

$$\epsilon_n = y_n - \hat{y}_n \quad \text{for} \quad n = 1, \ldots, N. \tag{2.21}$$

To evaluate generalization, the data are split into a training set $\mathcal{D}_{\text{train}}$ used to fit the parameters $\boldsymbol{\theta}$, and a test set $\mathcal{D}_{\text{test}}$ used to assess out-of-sample performance.

2.2.2 Bias–variance trade-off

To evaluate model performance and understand the trade-off between underfitting and overfitting, it is useful to decompose the expected test error into its constituent components. Recall from equation (2.19) that the observations follow the process $y_n = f(\mathbf{x}_n) + \epsilon_n$

with $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$. Suppose we have estimated a model $\hat{f}(\mathbf{x_n})$, and want to assess its accuracy on a new test point $(\mathbf{x}_t, y_t)$ which can be done through the expected squared prediction error, also known as the mean squared error (MSE) and given by

$$\text{MSE} = \mathbb{E}\left[\left(\hat{f}(\mathbf{x}_t) - y_t\right)^2\right]. \tag{2.22}$$

This term can be decomposed as

$$\mathbb{E}\left[\left(\hat{f}(\mathbf{x}_t) - y_t\right)^2\right] = \sigma^2 + \text{Var}\left[\hat{f}(\mathbf{x}_t)\right] + \left(f(\mathbf{x}_t) - \mathbb{E}\left[\hat{f}(\mathbf{x}_t)\right]\right)^2, \tag{2.23}$$

which is the bias–variance decomposition.

The first term, σ^2, the irreducible error, is noise in the data that cannot be explained by any model. The second term, $\text{Var}[\hat{f}(\mathbf{x}_t)]$, measures how much the model's variance in the estimation of the model and increases with model complexity. The third term is the squared bias, capturing the distance between the expected prediction and the true value, and typically decreases as model complexity increases.

In short, simpler models tend to underfit (high bias, low variance), while more complex models risk overfitting (low bias, high variance). Model selection aims to balance these effects, typically using cross-validation, regularization techniques, or ensemble methods to minimize prediction error and improve out-of-sample performance. Cross-validation provides a standard approach to balance bias and variance when assessing model generalization [24].

2.2.3 Linear models (OLS, Lasso)

Linear models assume a linear relationship between the dependent variable y_n and the independent variables $\mathbf{x}_n \in \mathbb{R}^p$ [23]. Formally, we write

$$y_n = \beta_0 + \boldsymbol{\beta} \cdot \mathbf{x}_n + \varepsilon_n \quad \text{where} \quad \varepsilon_n \sim N(0, \sigma^2), \tag{2.24}$$

where β_0 is the intercept, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p$ are the regression coefficients. We will denote by $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}$ the full parameter vector. The coefficients measure the sensitivity of implied volatility y_n for a given swaption structure on day n to the macroeconomic factors $\mathbf{x}_n$.

Ordinary least squares

The goal of ordinary least squares (OLS) is to find the parameter vector $(\beta_0, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}$ that minimizes the residual sum of squares (RSS), which is defined as the squared difference between observed and predicted values [23]. We will denote this vector by $\hat{\boldsymbol{\theta}}$, which formally solves the following minimization problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \ \mathrm{RSS}(\boldsymbol{\theta}) \qquad \text{where} \quad \mathrm{RSS}(\boldsymbol{\theta}) = \mathrm{RSS}(\beta_0, \boldsymbol{\beta}) = \sum_{n=1}^{N} \big(y_n - (\beta_0 + \boldsymbol{\beta} \cdot \mathbf{x}_n)\big)^2. \tag{2.25}$$

We can equivalently write

$$\mathrm{RSS}(\boldsymbol{\theta}) = \sum_{n=1}^{N} \left(y_n - \left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{n,j} \right) \right)^2 \tag{2.26}$$

$$= \sum_{n=1}^{N} \left(y_n - \sum_{j=0}^{p} \beta_j x_{n,j} \right)^2, \tag{2.27}$$

where we define $x_{n,0} := 1$ so that the intercept β_0 is included in the summation.

Thus, the OLS estimation problem can be written as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \left(y_n - \sum_{j=0}^{p} \theta_j x_{n,j} \right)^2. \tag{2.28}$$

Let $X \in \mathbb{R}^{N \times (p+1)}$ denote the design matrix whose n-th row is $(1, x_{n,1}, \ldots, x_{n,p})$, and let $y \in \mathbb{R}^N$ be the vector of responses. Then the OLS minimization problem becomes

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|X\boldsymbol{\theta} - y\|^2, \tag{2.29}$$

whose closed-form solution is

$$\hat{\boldsymbol{\theta}} = (X^\top X)^{-1} X^\top y, \tag{2.30}$$

assuming $X^\top X$ is invertible.

Lasso regularization

Although OLS minimizes the residual sum of squares, it can lead to overfitting when the model is too flexible or the features are highly correlated. Regularization is a statistical method to reduce errors caused by overfitting training data. To address this, we add a penalty term to the OLS objective that discourages large coefficients. The resulting framework is known as regularized least squares, where we solve

$$\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \left(y_n - \sum_{j=0}^{p} \theta_j x_{n,j} \right)^2 + \lambda \|\boldsymbol{\theta}_{1:p}\|_q^q, \tag{2.31}$$

with $\|\boldsymbol{\theta}_{1:p}\|_q^q = \sum_{j=1}^{p} |\theta_j|^q$, where $\|\cdot\|_q$ denotes the ℓ_q norm of the coefficient vector excluding the intercept. Here $\lambda > 0$ controls the strength of regularization.

The choice of q determines the type of regularization. If $q = 2$, we obtain Ridge regression, which shrinks all coefficients but keeps them in the model. If $q = 1$, we obtain the Lasso regression [25, 26, 23], which can shrink some coefficients exactly to zero, thereby performing variable selection in addition to shrinkage.

In this thesis, we focus on the Lasso estimator. The corresponding optimization problem becomes

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \left(y_n - \sum_{j=0}^{p} \theta_j x_{n,j} \right)^2 + \lambda \sum_{j=1}^{p} |\theta_j|, \tag{2.32}$$

where, as before, $x_{n,0} := 1$ ensures that the intercept θ_0 is not penalized.

Lasso is particularly useful in settings where only a subset of features is expected to drive the target. By eliminating irrelevant predictors, it can improve model interpretability and generalization. In our setting, Lasso performs variable selection among macroeconomic factors, helping identify which ones are most relevant for volatility.

2.2.4 Nonlinear models (Random forest)

Linear models are easy to interpret, but often too rigid to capture complex patterns. To address this, we also explore nonlinear models in the second stage of our framework. These models approximate the mapping $f(\mathbf{x}_n)$ without assuming a fixed functional form, which allows greater flexibility.

Random forest

A decision tree splits the input space into regions and makes predictions based on the average outcome in each region. It learns these splits by minimizing a loss function such as the residual sum of squares (RSS) in regression settings.

A random forest [27] is an ensemble of many such trees, each trained on a slightly different version of the data. Specifically, each tree is fitted to a bootstrapped sample, which means that it is trained on a random subset of the training data selected with replacement. As a result, each tree sees a slightly different view of the data. In addition, at each split within a tree, only a random subset of the features is considered. This further decorrelates the trees and improves generalization. The final prediction $\hat{y}_n$ is the average of all trees B

$$\hat{y}_n = \frac{1}{B} \sum_{b=1}^{B} T_b(\mathbf{x}_n), \tag{2.33}$$

where $T_b(\mathbf{x}_n)$ is the prediction of the b-th tree for $\mathbf{x}_n$.

By aggregating the predictions of many different trees, random forests reduce variance and often improve out-of-sample performance. The importance of variables can also be assessed by tracking how much each feature contributes to reducing the prediction error in the forest.

2.2.5 Feature selection methods (LOFO)

Feature selection methods are widely used in machine learning to identify which variables contribute most to predictive performance. Beyond regularization (as in Lasso) and ensemble-based importance measures (as in random forest), another approach is Leave one feature out (LOFO) analysis. LOFO provides a direct measure of each variable's marginal contribution by comparing model performance with and without that feature [28, 29]. The first formal description of the method was given by Lei et al. (2018), who introduced it under the name Leave one covariate out (LOCO) in the context of distribution-free inference [30].

Leave one feature out (LOFO)

Let $\hat{f}(\mathbf{x}; \hat{\boldsymbol{\theta}})$ denote a model with parameters $\hat{\boldsymbol{\theta}}$ fitted on the training set. We first evaluate the baseline error on the test set as

$$e_{\text{orig}} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} L\big(y_{\text{test}}^{(i)}, \hat{f}(\mathbf{x}_{\text{test}}^{(i)})\big), \tag{2.34}$$

where $L(\cdot, \cdot)$ is the mean squared error (MSE).

For each feature $j \in \{1, \dots, p\}$, we create modified training and test sets with the j-th feature removed, denoted $\mathbf{X}_{\text{train},-j}$ and $\mathbf{X}_{\text{test},-j}$. A new model $\hat{f}_{-j}$ is trained on the reduced feature set, and the error is recomputed as

$$e_{-j} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} L\big(y_{\text{test}}^{(i)}, \hat{f}_{-j}(\mathbf{x}_{\text{test},-j}^{(i)})\big). \tag{2.35}$$

The LOFO importance of feature j is defined as the difference between the error of the reduced model (without feature j) and the error of the original model:

$$\text{LOFO}_j = \varepsilon_{\text{without } j} - \varepsilon_{\text{orig}}. \tag{2.36}$$

Features with larger positive LOFO_j values are more important, since removing them causes a greater deterioration in predictive performance. Negative values of LOFO_j indicate that the error decreases when feature j is removed, suggesting that the variable may add noise rather than explanatory power.

2.3 Evaluation metrics

To evaluate the performance of the models introduced in the second stage of our two-stage framework, we assess their ability to predict the deviations of implied volatility from the fitted backbone. The same setup described in the previous section applies here.

We consider inputs $\mathbf{x}_n \in \mathbb{R}^p$, targets $y_n \in \mathbb{R}$, and model predictions $\hat{y}_n = f(\mathbf{x}_n; \theta)$. We compute prediction errors on a training set $\mathcal{D}_{\text{train}}$ and a separate test set $\mathcal{D}_{\text{test}}$ to evaluate both fit and generalization.

Using this notation, we introduce the following metrics to assess accuracy and explanatory power: the mean squared error (MSE) and the coefficient of determination (R^2).

2.3.1 Mean squared error

A fundamental way to assess model performance is through the mean squared error (MSE). MSE measures how well a model fits the observed data and is defined as the average of the squared differences between the observed values y_n and the model predictions $\hat{y}_n$

$$\text{MSE} = \frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 . \tag{2.37}$$

When the model is evaluated on the same data used to fit it, i.e., the training data $\mathcal{D}_{\text{train}}$, we obtain the in-sample error. In practice, we are often more interested in the model's ability to generalize, i.e., its performance on unseen data. This is captured by computing the MSE on the test set $\mathcal{D}_{\text{test}}$

$$\text{MSE}_{\text{test}} = \frac{1}{N_{\text{test}}} \sum_{n=1}^{N_{\text{test}}} (y_n - \hat{y}_n)^2 . \tag{2.38}$$

A lower MSE, whether in-sample or out-of-sample, indicates that the model predictions are closer to the true values. Importantly, MSE is a general metric and can be used for both linear and non-linear models, since it does not rely on any assumption about the form of the model.

2.3.2 R-squared (coefficient of determination)

Another commonly used performance metric is the coefficient of determination, R^2, which measures the proportion of variance in the data that is explained by the model

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}, \tag{2.39}$$

where

$$\text{RSS} = \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 \tag{2.40}$$

is the residual sum of squares, and

$$\mathrm{TSS} = \sum_{n=1}^{N} \left(y_n - \bar{y}\right)^2, \quad \text{with} \quad \bar{y} = \frac{1}{N} \sum_{n=1}^{N} y_n, \tag{2.41}$$

is the total sum of squares around the mean. An R^2 value close to 1 indicates that the model explains most of the variance in the data, while an R^2 near 0 suggests limited explanatory power. In our framework, we compute R^2 on both training and test sets to evaluate fit and generalization. Although R^2 is most commonly interpreted in the context of linear models, many libraries also report an R^2 value for non-linear models as a descriptive statistic of explained variance.

2.3.3 Relationship between MSE and R-squared

Both metrics are linked through the residual sum of squares

$$\mathrm{RSS} = \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 = N \times \mathrm{MSE}. \tag{2.42}$$

A smaller MSE leads to a smaller RSS, which in turn increases R^2, provided the total sum of squares remains constant. Thus, in model comparison, improvements in MSE are typically reflected in improvements in R^2 as well.

In this thesis, we report both training and test values of MSE and R^2 to ensure that our models explain a meaningful proportion of the variance without overfitting and that they generalize well beyond the training data.

2.3.4 Adjusted R-squared

While R^2 increases (or at least does not decrease) as more predictors are added, it does not account for model complexity. To address this, the adjusted coefficient of determination, adjusted R^2, introduces a penalty for the number of predictors relative to the sample size

$$R^2_{\mathrm{adj}} = 1 - \left(1 - R^2\right) \frac{N-1}{N-p-1}, \tag{2.43}$$

where N is the number of observations and p the number of predictors. Adjusted R^2 therefore only increases when a new predictor improves model fit. In our framework, we report adjusted R^2 on the test set to enable fairer comparisons between models with different numbers of features.

2.3.5 Pearson correlation coefficient

The Pearson correlation coefficient [31] assesses the linear association between the modeled and observed data. It is computed as

$$\rho = \frac{\mathrm{Cov}(y_t, \hat{y}_t)}{\sigma_y \sigma_{\hat{y}}}, \tag{2.44}$$

where $\mathrm{Cov}(\cdot)$ is the covariance and σ_y, $\sigma_{\hat{y}}$ are the standard deviations of the observed and modeled data. The covariance is defined as

$$\mathrm{Cov}(y_t, \hat{y}_t) = \frac{1}{T-1} \sum_{t=1}^{T} (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}}), \tag{2.45}$$

with $\bar{y}$ and $\bar{\hat{y}}$ denoting the sample means. A value of ρ close to 1 indicates a strong positive linear relationship, while a value near 0 suggests no linear relationship. A value close to -1 indicates a strong negative linear relationship.

3 Methodology

This chapter describes the methodology used in the thesis. We begin by introducing the data sources, including the swaption implied volatilities and the macroeconomic factors. We then set up the modeling framework consisting of two stages. Stage 1 fits a structural volatility–rate relationship that captures the dependence of implied volatility on forward yields. Stage 2 explains the deviations from this backbone using macroeconomic drivers. Finally, we present the machine learning models used in Stage 2 and the evaluation metrics applied to assess statistical performance.

3.1 Data sources

All market and macroeconomic data used in this thesis were retrieved from the Bloomberg Terminal and carefully pre-processed to ensure consistency across all series.

The swaption data set consists of daily observations of at-the-money implied volatilities for a set of expiry–tenor combinations. Specifically, it covers six option expiries (3 months, 6 months, 1 year, 2 years, 3 years, and 5 years) and seven underlying swap tenors (1 year, 2 years, 3 years, 5 years, 7 years, 10 years, and 30 years), forming a 6×7 grid across the swaption volatility surface. Here, each combination refers to a specific pairing of an option expiry and an underlying swap tenor (e.g., a 1y$\times$10y swaption is a 1-year option on a 10-year swap). The full set of such combinations across all expiries and tenors is often referred to as the swaption volatility cube, since implied volatilities can be visualized in three dimensions: expiry, tenor, and volatility level. For each pair of expiry-tenors, we collect the corresponding forward swap rate and the associated implied volatility in two formats: basis point volatility and yield volatility.

Figure 3.1 shows the mean and standard deviation of yield implied volatilities across the expiry–tenor grids for USD and EUR. In both markets, volatility declines with longer option and swap maturities. This is consistent with lower sensitivity at the long end of the curve. The plots also show that short-dated swaptions have both higher average volatility and more variability over time, with this effect being stronger in EUR. These surfaces illustrate the main features of the volatility cube and show that there are structural differences between the USD and EUR markets.

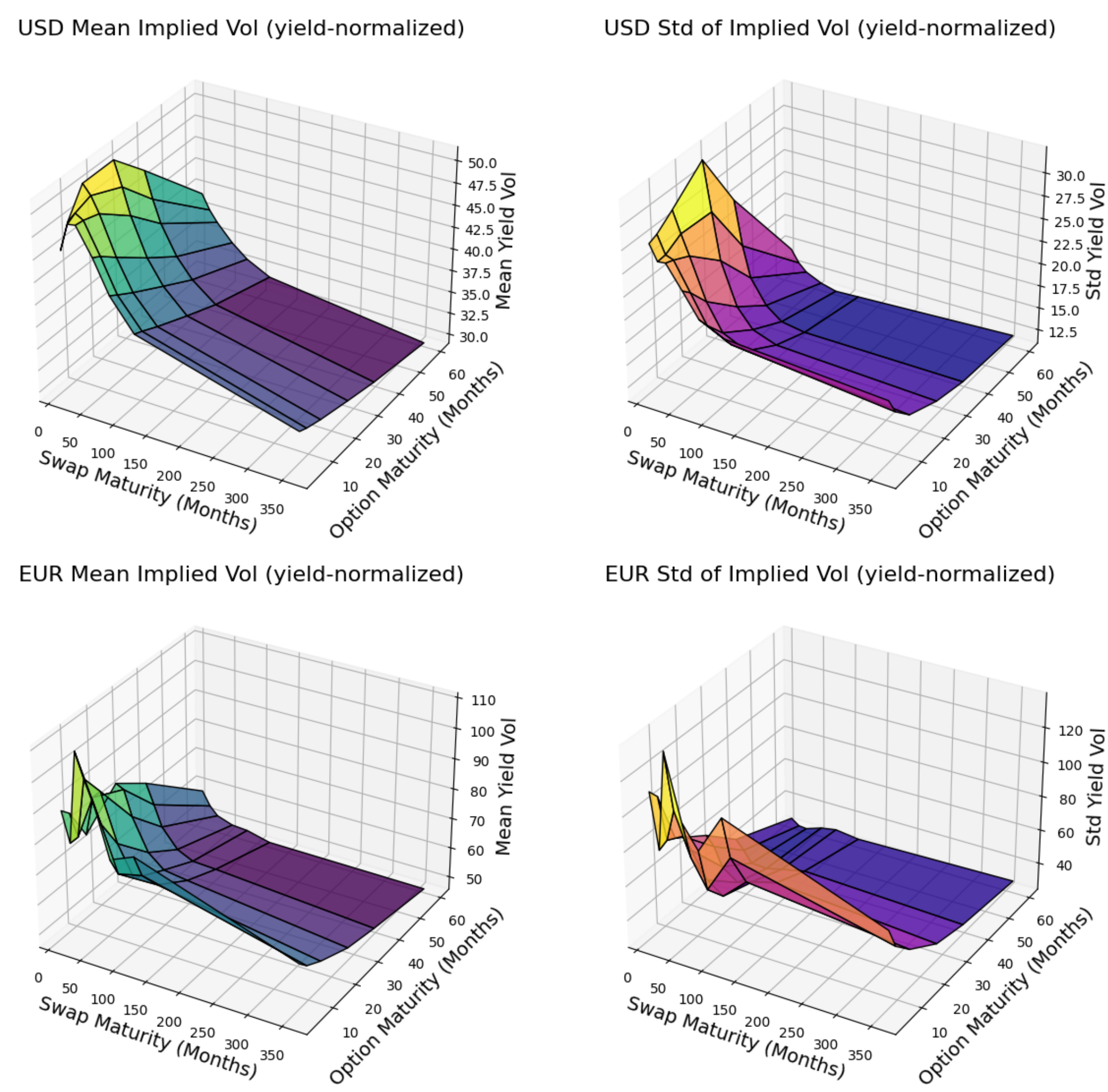

Figure 3.1: Mean and standard deviation of yield implied volatilities across the USD and EUR swaption cubes. Both markets show higher volatility at the short end that declines with maturity, while EUR displays greater variability and less stability across the surface.

The historical swaption data set spans the period from 2 January 2014 to 30 May 2025, providing over 11 years of daily data. This allows us to capture a variety of monetary policy regimes, including low-rate environments, quantitative easing, and recent post-pandemic behaviors.

In addition to market data, we incorporate macroeconomic variables that are expected to influence swaption volatility. Specifically, we focus on five factors commonly used in both academic and industry models: the current forward rate, 5yx5y inflation swap, central bank balance sheet size, 10-year swap spread, and 3mx3m–1yx3m curve slope. These variables serve as proxies for rate levels, inflation expectations, liquidity provision, market funding conditions, and policy expectations.

The macroeconomic data span from early January 2000 or January 2004, depending on availability, to the end of May 2025. These factors are discussed in detail in Section 3.4.1, where we analyze their role in explaining deviations from the structural volatility–rate relationship. For USD regressions, the balance sheet variable refers to the Federal Reserve System (Fed), while for EUR regressions it refers to the European Central Bank (ECB). Any missing observations, such as weekly balance sheet data, were interpolated to ensure alignment with the daily swaptions data (see Section 3.2 for details).

These five variables form a 5-dimensional macro factor explanatory vector x_n used to model deviations from the volatility backbone. The corresponding target variable y_n is the swaption implied volatility for a given expiry–tenor on day n.

The structure of this data set closely follows that used in JP Morgan's two-stage volatility modeling framework [5, 6]. While JP Morgan's implementation focused on USD markets, the inclusion of EUR in this thesis serves as an extension to test whether the framework remains valid in different currency regimes.

3.2 Data processing

Once retrieved from Bloomberg, all series were carefully aligned and pre-processed before being used in the modeling framework. The swaption data, underlying swap rates, and macroeconomic factors were originally stored in separate data sets with differing frequencies. To enable joint analysis, all series were merged onto a common daily timeline. The swaption data set contains at-the-money implied volatilities across the full expiry–tenor grid, reported in both basis point and yield terms. To merge this data with macroeconomic factors, all sources were synchronised to daily frequency.

A key step in this process was the treatment of variables originally reported at weekly frequency. Specifically, balance sheet data from the Federal Reserve System (Fed) and the European Central Bank (ECB) were interpolated with cubic splines to produce smooth daily series aligned with the swaption data. This approach preserves the underlying trend without introducing artificial jumps. Figure 3.2 illustrates the original weekly USD balance sheet values (blue) alongside the interpolated daily series (orange). This ensures that the macro factor vector x_n is available for each trading day.

After interpolation, all time series were validated for completeness and consistency. Finally, the series were merged by date into a unified data set, ensuring that for each day n the explanatory vector x_n, comprising the current forward rate, 5yx5y inflation swap, central bank balance sheet size, 10-year swap spread, and 3mx3m–1yx3m curve slope, was fully aligned with the corresponding target variable y_n, the observed swaption implied volatility for a given expiry–tenor pair.

This preprocessing step ensures that subsequent modeling stages are based on a consistent and complete data set, allowing the two-stage framework to be implemented without additional adjustments.

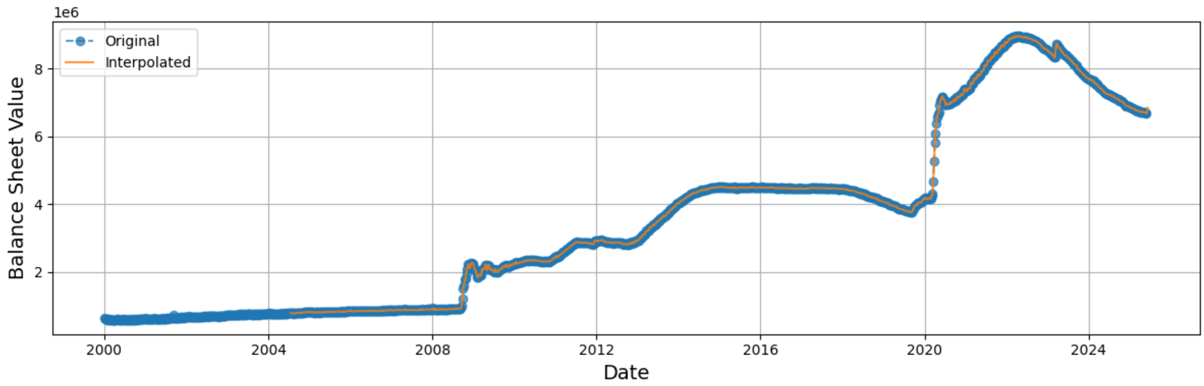

Figure 3.2: USD balance sheet series. Weekly observations (blue) are interpolated with cubic splines to obtain a smooth daily series (orange) consistent with the frequency of the swaption data.

3.3 Stage 1: Structural volatility fit

The first stage of our framework establishes a cross-sectional volatility–rate relationship, or 'backbone', that captures the fundamental dependence of implied volatility on the level of underlying forward rates. This step follows the approach introduced in JP Morgan's two-stage framework [5, 6] and aligns with the empirical findings of Rebonato et al. [2]. By fitting this structural relationship first, we remove the dominant level effect of interest rates on volatility so that, in Stage 2, we can focus on deviations driven by macroeconomic factors.

Empirical studies show that implied volatility is not constant across rate levels. Instead, it follows a three-regime pattern: volatility tends to scale proportionally with the rate at very low levels, flattens out in a mid-range, and increases again at high levels. This 'universal relationship' has been documented across currencies and tenors [2]. Capturing this effect in a smooth way provides a stable baseline for further analysis.

We fit a single exponential function to model yield volatility across all swaption observations, combining daily data across all expiries and tenors. We use a single exponential fit rather than piecewise, as it captures the overall relationship and gives a stable backbone across the surface. We model the percent annualized yield volatility as an exponential function of the forward yield,

$$\sigma_{\text{yld}}(y) = a + b\exp(-cy), \tag{3.1}$$

where y is the at-the-money forward yield and $\sigma_{\text{yld}}(y)$ is the fitted percent yield volatility. This form allows the normal-vol-versus-rate curve to exhibit a low slope in the middle and higher slopes at low as well as high yield levels [5, 6]. Using the exponential function, the parameters a, b and c were estimated via nonlinear least squares by minimizing the squared error between observed and modeled yield volatilities. This was done using the `curve_fit` function from the `scipy.optimize` module.

Once the parameters were estimated, the fitted percent-yield volatilities were translated into a normal basis point volatility backbone using

$$\sigma_{\text{bp}}(y) = \big[a + b\exp(-cy)\big]\frac{y}{\sqrt{251}}, \tag{3.2}$$

where the conversion from yield volatility to basis point volatility is given by

$$\sigma_{\text{bp}} = \sigma_{\text{yld}} \times y, \tag{3.3}$$

and the division by $\sqrt{251}$ converts annualized basis point volatility into daily units, matching the daily frequency of the swaptions in our data. In equation 3.1 and 3.2, $\sigma_{\text{bp}}(y)$ denotes daily basis point volatility, while $\sigma_{\text{yld}}(y)$ in denotes annualized percent yield volatility.

The final step is to explain why we divide by $\sqrt{251}$ to obtain daily volatility. For notational simplicity, in this derivation we write $\sigma(y)$ to denote annualized basis point volatility. At the end we will distinguish between annualized volatility $\sigma(y)$ and daily volatility $\sigma_{\text{daily}}(y)$. This adjustment comes from the Black–Scholes [17] convention of defining volatility as the annualized standard deviation of returns. In a diffusion model,

$$\frac{dS_t}{S_t} = \mu dt + \sigma(y)\, dW_t, \tag{3.4}$$

the log-return process is given by

$$r_t = \log\frac{S_t}{S_0} = \big(\mu - \tfrac{1}{2}\sigma^2(y)\big)\, t + \sigma(y)W_t, \tag{3.5}$$

where W_t is a Wiener process with variance $\text{Var}(W_T) = T$, $\sigma(y)$ denotes the annualized basis point volatility backbone, and T is measured in years.

It follows that the variance of returns grows linearly with time,

$$\text{Var}(r_T) = \text{Var}\big(\big(\mu - \tfrac{1}{2}\sigma^2(y)\big)\, T + \sigma(y)W_T\big) = \text{Var}(\sigma(y)W_T) = \sigma^2(y)\,\text{Var}(W_T) = \sigma^2(y)\, T. \tag{3.6}$$

For a single trading day, we set $T = \frac{1}{251}$, which gives

$$\text{Var}(r_{\text{daily}}) = \sigma^2(y) \cdot \tfrac{1}{251}. \tag{3.7}$$

Daily volatility is then defined as the standard deviation of one-day returns,

$$\sigma_{\text{daily}}(y) = \sqrt{\text{Var}(r_{daily})} = \sqrt{\sigma^2(y) \cdot \tfrac{1}{251}} = \frac{\sigma(y)}{\sqrt{251}}. \tag{3.8}$$

Hence, daily volatility is obtained from annualized volatility by

$$\sigma_{\text{daily}}(y) = \frac{\sigma(y)}{\sqrt{251}}. \tag{3.9}$$

Thus, dividing by $\sqrt{251}$ ensures that the fitted backbone $\sigma(y)$ is expressed in daily volatility units, matching the frequency of the observed swaption data.

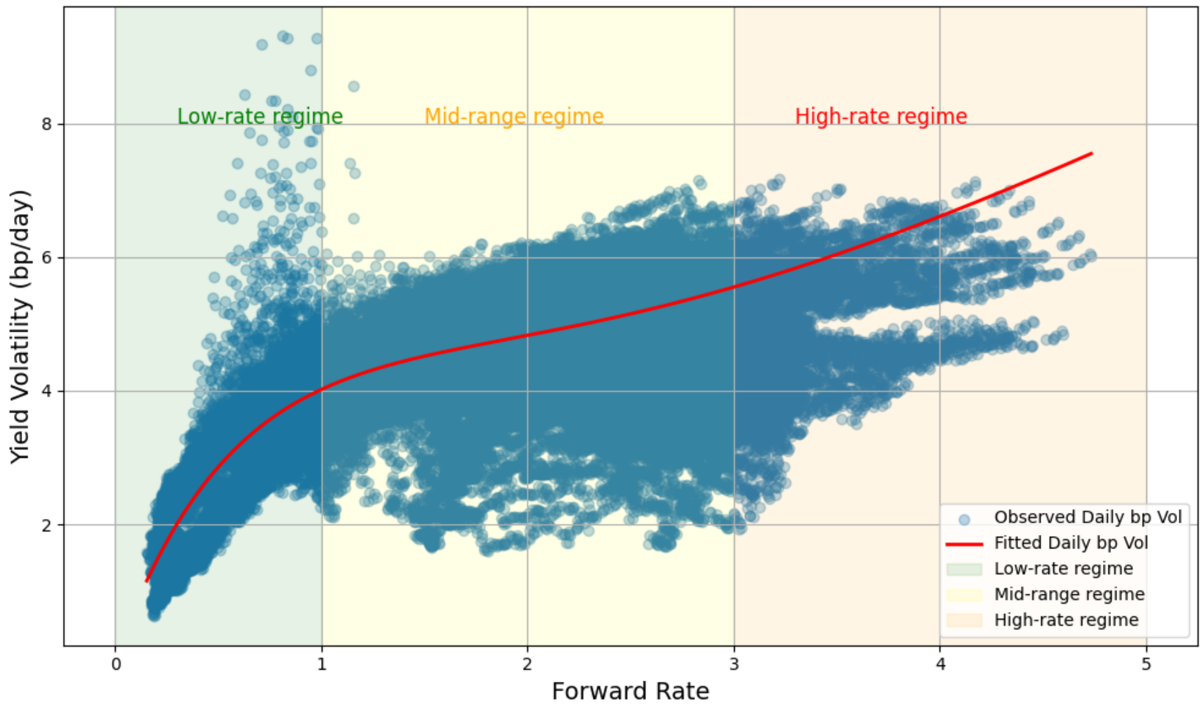

Figure 3.3: Observed daily basis point volatilities (blue) versus fitted volatility backbone (red) for USD. The fitted curve captures the three-regime pattern documented in the literature, with steeper slopes at low and high yields and a flatter section in the middle.

Figure 3.3 illustrates the fitted relationship using all available data for USD in our data set. The blue dots show the observed daily basis point volatilities plotted against their corresponding forward yields across the entire sample, while the red line shows the fitted backbone curve. The fitted curve reproduces the three regime pattern reported in the literature, with higher slopes at low and high yields and a flatter region in between. This agreement with the findings of Rebonato et al. [2] provides confidence that the fitted backbone is a valid structural representation of the volatility–rate relationship. The corresponding plot for EUR (not shown) displays a very similar shape, confirming that the backbone relationship holds consistently across both markets, which is in line with the literature [1] [2].

The Stage 1 fit captures the universal relationship between rates and volatility, providing a robust baseline. In Stage 2, we measure daily deviations from this backbone and explain them using macroeconomic factors.

3.4 Stage 2: Modeling deviations with macroeconomic factors

The second stage of the framework addresses the residual component of implied volatility after removing the dominant volatility–rate dependence captured by the Stage 1 backbone. While the backbone reflects the universal link between forward rates and volatility [1, 2], implied volatility is also shaped by broader macro-financial conditions [6]. We therefore model these deviations as a function of five macroeconomic variables.

3.4.1 Selection and role of macroeconomic factors

Figure 3.4 plots observed implied volatilities against each of the five macro factors for representative expiry–tenor combinations. Each column corresponds to a factor, and each row to a swaption structure. This provides a first impression of how volatility moves with these variables. Each factor is introduced below, together with the proxy used in the model and its relationship to volatility.

Forward rates. The dependence of volatility on rate levels is a well-documented 'universal relationship' across currencies and maturities [1]. Although this dependence is primarily captured in Stage 1, forward rates are retained in Stage 2 to control for sector-specific deviations. In Figure 3.4, the forward rate axis shows a clear positive slope for most combinations. This confirms a systematic relationship between volatility residuals and the underlying rate level.

Long-term inflation expectations. Inflation uncertainty has been identified as a key driver of short-rate volatility over the past three decades [3]. We proxy long-term expectations with the 5yx5y inflation swap. This is a market instrument that reflects expected inflation over a five-year period beginning five years ahead. In Figure 3.4, volatility is lowest when the 5yx5y inflation swap is close to 2%, consistent with the idea that credible inflation targets stabilize rates, while deviations in either direction are associated with higher volatility.

Central bank balance sheet size. The size of central bank balance sheets (Fed for USD, ECB for EUR) captures the extent of quantitative easing (QE) or quantitative tightening (QT) and overall market liquidity. Larger balance sheets are typically associated with lower volatility, reflecting liquidity provision. This negative relation is visible in Figure 3.4, where volatility tends to fall as balance sheet size increases.

Swap spread. The 10-year swap spread, defined as the difference between the 10-year swap rate and the 10-year government bond yield, serves as a proxy for funding costs and credit premia. Wide or unstable spreads often indicate market stress and are typically associated with higher volatility. Figure 3.4 shows several expiry–tenor combinations where volatilities increase with wider spreads.

Curve slope. The slope between the 3m3m and 1y3m forward rates captures near-term monetary policy expectations. A steep slope signals anticipated policy shifts, while a flat curve indicates stable expectations. In Figure 3.4, changes in slope are associated with clear shifts in volatility.

Together, these five variables serve as proxies for the main macroeconomic forces, including rate levels, inflation expectations, liquidity, funding conditions, and policy outlook, that help explain deviations from the structural volatility–rate relationship. They form the explanatory vector $\mathbf{x}_n$ used in the second stage.

3.4.2 Model formulation

The fitted volatility backbone from Stage 1 is combined with the macro factor deviations in a single expression

$$\sigma_{\mathrm{bp}}(y) = (a + be^{-cy_n})\frac{y_n}{\sqrt{251}} + \beta_1 x_{n,1} + \beta_2 x_{n,2} + \cdots + \beta_5 x_{n,5}, \tag{3.10}$$

where the first term reproduces the daily basis point volatility backbone and the second term captures the modeled deviation as a linear combination of macro factors. The coefficients β_j are estimated individually for each expiry–tenor combination using the residuals from Stage 1 as the target variable.

This two-stage approach, with a universal backbone and a structured macro factor adjustment, provides a robust framework for modeling implied volatility dynamics across USD and EUR markets. In Section 4, we analyze the estimated coefficients and compare them with economic intuition, using the scatter plots above as a reference point.

3.5 Machine learning methods for volatility modeling

In the second stage of our framework we model the deviations of implied volatility from the fitted volatility backbone obtained in Section 3.3. For each trading day n, the explanatory vector x_n contains the five macroeconomic factors, and the target variable y_n is the residual from Stage 1, i.e. the difference between the observed daily basis point volatility and the fitted structural volatility. By construction, these residuals represent the portion of volatility not explained by the universal rate–volatility relationship, allowing us to focus on additional drivers.

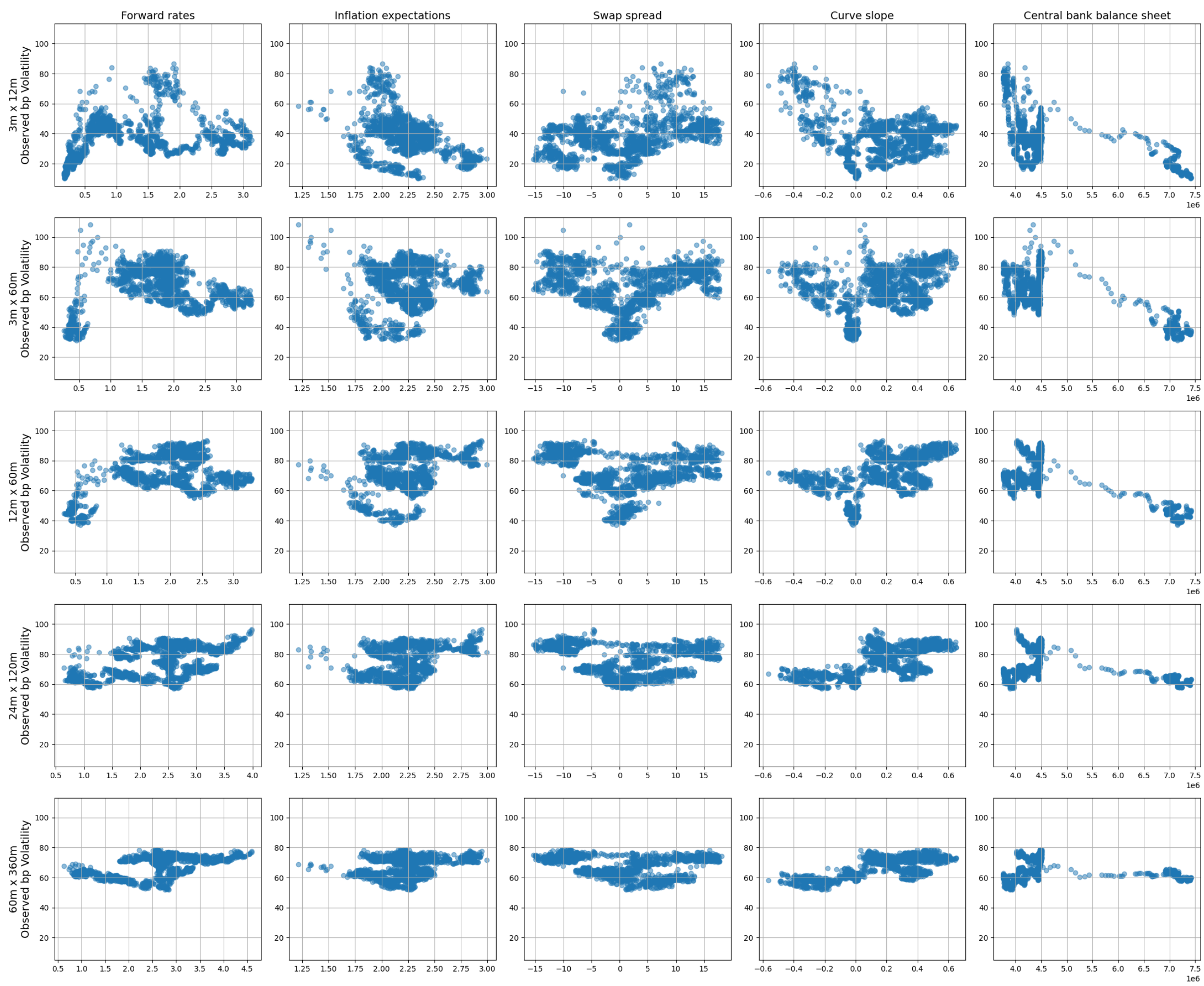

Figure 3.4: Scatter plots of observed implied volatilities against the five selected macroeconomic factors for a representative set of expiry–tenor combinations for USD. Each column corresponds to a macro factor and each row to an expiry–tenor combination. Forward rates show a clear positive slope, inflation swaps point to lowest volatility near the 2% target, and balance sheet size is negatively related to volatility. Swap spreads are associated with higher volatility when spreads widen, while shifts in the curve slope correspond to changes in volatility across the cube.

We split the data set into a training set $\mathcal{D}_{\text{train}}$ and a test set $\mathcal{D}_{\text{test}}$. Models are trained on $\mathcal{D}_{\text{train}}$ and evaluated on $\mathcal{D}_{\text{test}}$ to ensure that predictive performance is not overestimated due to overfitting.

We consider both linear and non-linear models, chosen to capture potentially different types of dependencies between macro factors and volatility residuals.

Ordinary Least Squares (OLS). OLS serves as our baseline. It assumes a linear relationship between residual volatility and macro factors and provides interpretable coefficients that can be compared against economic intuition.

Leave One Feature Out (LOFO). LOFO re-estimates the regression repeatedly, each time excluding one of the macro factors. Then it compares performance across these reduced models and informs which variables contribute most to explanatory power.

Lasso regression. Lasso adds an L_1 penalty to the regression coefficients, enabling automatic feature selection by shrinking less relevant coefficients to zero.

Random forest. A tree-based ensemble method that builds multiple decision trees on bootstrapped samples and averages their predictions. In our framework, it is used both to capture potential nonlinear relationships and to identify the relative importance of each macro factor through feature selection.

The results in Chapter 4 will show how these models compare and which macro drivers play the most significant role in explaining volatility deviations.

3.6 Model evaluation and validation

To assess the quality of the fitted models we use the metrics introduced in 2.3, namely the mean squared error (MSE), the coefficient of determination R^2, and the adjusted R^2. MSE provides a direct measure of the average squared prediction error, while R^2 indicates the proportion of variance in the residual volatility explained by the model. Adjusted R^2 further accounts for model complexity, penalising the inclusion of irrelevant predictors.

We report both in-sample (training) and out-of-sample (test) values to monitor overfitting. A large discrepancy between training and test performance would indicate that the model captures noise rather than underlying patterns. In addition, we interpret our results through the lens of the bias–variance trade-off.

Beyond these standard metrics, we also assess robustness and generality. First, we check cross-currency consistency by verifying that results obtained for USD extend to EUR. Second, we conduct sensitivity analyses of the Stage 2 coefficients (OLS LOFO, Lasso-selected OLS, and random forest feature importance) to identify which macro factors drive volatility and whether their signs align with economic intuition. This combined evaluation ensures that the selected model not only generalizes statistically but also provides economically interpretable insights into the drivers of swaption volatility.

4 Results

Before turning to the individual models, we first explain the regression setup used throughout this chapter. The goal is to study how well macroeconomic factors can explain movements in implied volatility. For each expiry–tenor combination of the USD and EUR swaption cubes we run a separate regression of implied volatility residuals on macroeconomic drivers. This results in 42 regressions per currency.

The dependent variable is the volatility residual, while the independent variables, or predictors, are five macroeconomic factors: the current forward rate, long-term inflation expectations (5yx5y inflation swap), the size of the central bank balance sheet, the 10-year swap spread, and the short-term curve slope (3mx3m–1yx3m). These factors were introduced and motivated in Section 3.4.1. Throughout this chapter, we will use the shortened terms forward rate, inflation swap, balance sheet, swap spread, and curve slope for simplicity.

The data is split into a training and test set using an 80/20 random split. Our main focus is on the full sample from 2014 to 2024. This window is chosen because it spans several distinct interest rate environments: first, an extended period of very low rates after the financial crisis; then a gradual move back toward more normal levels before the pandemic; and finally, the sharp hikes during the recent inflation period. Using the full sample allows us to capture how the relationship between volatility and macroeconomic drivers behaves across these different regimes.

In the following sections we first fit OLS using all five macro factors. We then test whether model performance can be improved through feature selection, applying three approaches: a leave one variable out procedure, the regularization method Lasso, and the nonlinear model random forest. The selected features from each method are then used to re-fit OLS models, and all models are evaluated on the same 80/20 split to ensure comparability. This full window spans distinct regimes and is our baseline for comparing model stability across time.

4.1 OLS with all five factors

We begin with the baseline model where all five macro factors are included simultaneously in the OLS regression. Over the full 2014–2024 sample, the mean out-of-sample R^2 is

0.48 for USD swaptions and 0.67 for EUR swaptions. This suggests that the linear model provides a reasonable fit, but still leaves a significant portion of the variation unexplained.

As a first step we test whether the estimated coefficients are statistically significant. This is done using the standard t-test, where the null hypothesis is that the coefficient equals zero. If the corresponding p-value is below 0.05, we reject the null hypothesis $H_0 : \beta = 0$ at the 5% level, indicating that the estimated coefficient is statistically different from zero. Tables 4.1 and 4.2 summarize the results for all expiry–tenor combinations.

Table 4.1: Summary of significance of predictors across expiry×tenor combinations for USD. Forward rate and balance sheet variables are significant in all regressions, while the inflation swap, curve slope, and the swap spread are significant in most cases. Non-significant results are concentrated in a few medium- and long-tenor combinations.

Macro factor	**Significant (n/total (%))**	**Non-significant combinations**
Forward rate	42/42 (100%)	—
Inflation swap	36/42 (85.7%)	12m×24m, 12m×36m, 24m×24m, 24m×36m, 36m×12m, 36m×24m
Balance sheet	42/42 (100%)	—
Swap spread	34/42 (81.0%)	3m×60m, 3m×84m, 3m×120m, 6m×60m, 6m×84m, 12m×60m, 24m×36m, 36m×24m
Curve slope	36/42 (85.7%)	3m×60m, 3m×84m, 6m×60m, 6m×84m, 12m×60m, 24m×36m

Table 4.2: Summary of significance of predictors across expiry×tenor combinations for EUR. Forward rate is significant in all regressions, while the inflation swap, balance sheet, swap spread, and curve slope are significant in most cases.

Macro factor	**Significant (n/total (%))**	**Non-significant combinations**
Forward rate	42/42 (100%)	—
Inflation swap	39/42 (92.9%)	3m×120m, 6m×24m, 12m×12m
Bank balance sheet	38/42 (90.5%)	3m×120m, 6m×120m, 24m×84m, 36m×60m
Swap spread	39/42 (92.9%)	12m×24m, 12m×36m, 24m×12m
Curve slope	37/42 (88.1%)	3m×12m, 6m×12m, 24m×360m, 36m×360m, 60m×360m

For USD swaptions (Table 4.1), the forward rate and the balance sheet variable are significant in the 42 regressions. The inflation swap and the curve slope are significant in roughly 86% of cases, while the swap spread is significant in 81% of cases. Non-significant results are limited to a small number of medium- and long-tenor combinations. For EUR swaptions (Table 4.2), significance is even higher. The forward rate is again significant in all regressions, while all other factors exceed a 88% significance rate. Overall, this confirms that almost all coefficients are statistically different from zero, and we can safely reject the null hypothesis for the majority of expiry–tenor combinations.

Having established statistical significance, the next step is to examine the estimated coefficients. In linear regression, the model finds the combination of predictors that best explains the dependent variable. For each factor, the estimated coefficient indicates how a one unit increase in that factor changes implied volatility. A positive coefficient means that volatility increases when the factor rises, while a negative coefficient means that volatility decreases. Since the predictors are on very different scales, the absolute size of coefficients is not directly comparable across factors. For example, the balance sheet variable takes values in the hundreds of thousands, which produces very small coefficient estimates. For this reason, we focus on the signs of coefficients and their stability across expiry–tenor combinations, rather than their absolute magnitude.

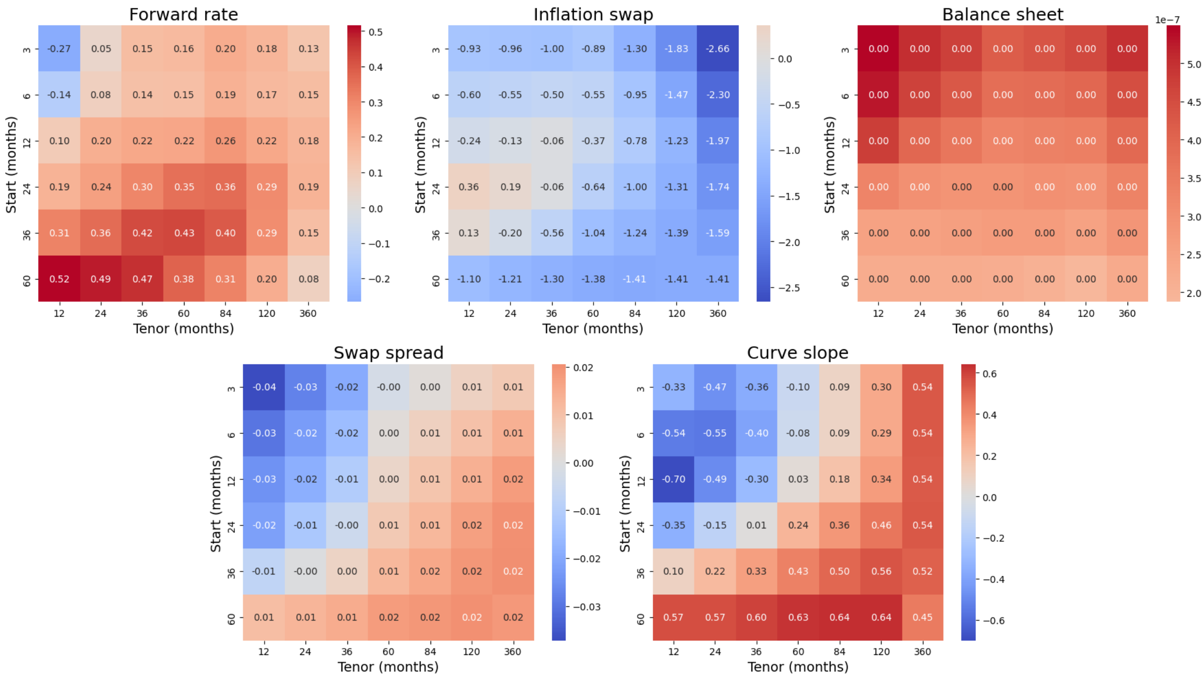

Figure 4.1: Estimated OLS coefficients for USD across expiry–tenor combinations. Each panel shows one macro factor, with colours indicating the sign and magnitude of the coefficient. The results show that forward rates and balance sheet effects generally enter with positive signs, while inflation swaps are negative across most of the surface. Swap spreads are close to zero or mildly negative, and the curve slope shifts from negative at the short end to strongly positive at longer maturities.

Figures 4.1 and 4.2 show the OLS coefficient heatmaps for USD and EUR. For USD, the forward rate enters with positive coefficients, which increase in size at longer expiries and tenors. The inflation swap is negative across almost all combinations, with larger magnitudes at the long end of the cube. The balance sheet coefficients are always positive and are higher for short expiries and tenors than elsewhere. The swap spread is slightly negative at the short end and closer to zero or mildly positive at longer maturities. The curve slope is negative at the short end but becomes strongly positive for longer expiries and tenors. For EUR, the sign structure differs notably. The forward rate is negative

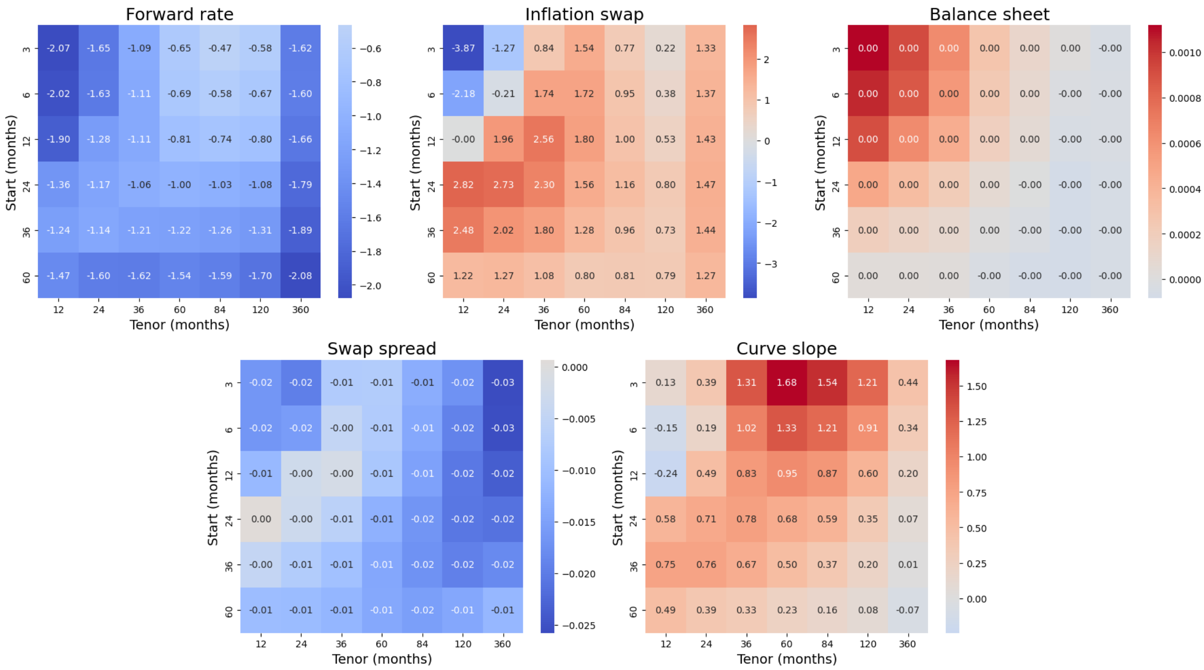

Figure 4.2: Estimated OLS coefficients for EUR across expiry–tenor combinations. Each panel shows one macro factor, with colours indicating the sign and magnitude of the coefficient. The results show that forward rates are negative across the surface, inflation swaps are mostly positive and strongest at medium expiries and tenors, and balance sheet effects are concentrated at the short end. Swap spreads are consistently negative, while the curve slope is positive throughout with the largest values at short expiries and medium tenors.

throughout, with larger values at short expiries and medium tenors, and at long tenors. The inflation swap is positive for most combinations, strongest at medium expiries and tenors. The balance sheet again shows positive coefficients with higher ones at short expiries and tenors. The swap spread is consistently negative across the surface, with higher values at high tenors. The curve slope is positive throughout, with the largest coefficients at short expiries and medium tenors.

Overall, the heatmaps show stable and interpretable sign patterns within each currency, but also highlight structural differences between them, which will be examined further in the discussion section.

Finally, we assess correlations. Table 4.3 reports the correlation between each macro factor and the residuals, based on 107,167 observations for USD and 86,194 observations for EUR. In our case, a positive correlation means that higher values of a macroeconomic factor are typically associated with higher volatility residuals, while a negative correlation implies that higher factor values are linked to lower residuals.

For USD, the balance sheet stands out with the strongest positive correlation ($\rho = 0.515$). The curve slope shows a moderate negative correlation ($\rho = -0.167$). For EUR, the strongest relationships are observed for the curve slope ($\rho = 0.535$), forward rate ($\rho = -0.409$), and inflation swap ($\rho = -0.309$). The swap spread plays only a minor role in USD ($\rho = -0.014$) but is more substantial in EUR ($\rho = -0.398$).

Table 4.3: Correlation of macro factors with residuals for USD and EUR. Pearson correlation coefficients ρ (rounded to three decimals). For USD, the residual volatility correlates most strongly with balance sheet size, while the other factors show weak correlations. For EUR, the forward rate, inflation swap, and swap spread are negatively correlated, whereas the curve slope shows a clear positive correlation.

Macro factor	$\rho_{\mathbf{USD}}$	$\rho_{\mathbf{EUR}}$
Forward rate	0.112	-0.409
Inflation swap	0.091	-0.309
Balance sheet	0.515	0.098
Swap spread	-0.014	-0.398
Curve slope	-0.167	0.535

Tables 4.4 and 4.5 show the correlations among macro factors. Again, a positive value means that the two factors tend to move together, while a negative value means that they usually move in opposite directions. In both currencies there are some notable relationships, such as a strong positive correlation between the forward rate and inflation swap, and negative correlations between the swap spread and the curve slope. This motivates a closer look at multicollinearity in later sections, although no extreme values are observed here.

Table 4.4: Correlation matrix among macroeconomic factors (USD). Pearson correlation coefficients ρ (rounded to three decimals). Forward rates and inflation swaps are strongly positively correlated, while both show weaker positive links with the balance sheet. Swap spreads are negatively correlated with inflation swaps and the curve slope, and the curve slope is negatively correlated with all other factors.

	FR	**IS**	**BS**	**SS**	**CS**
Forward rate (FR)	1.000	0.624	0.190	-0.139	-0.342
Inflation swap (IS)	0.624	1.000	0.344	-0.474	-0.062
Balance sheet (BS)	0.190	0.344	1.000	-0.129	-0.322
Swap spread (SS)	-0.139	-0.474	-0.129	1.000	-0.344
Curve slope (CS)	-0.342	-0.062	-0.322	-0.344	1.000

To summarize, OLS with all five factors achieves moderate explanatory power, with better fit for EUR than USD. Almost all coefficients are statistically significant, and the signs are generally consistent across the swaption cube, though they differ between currencies. Given the length and complexity of the 2014–2024 period, the model may underfit. In the next step we examine whether model performance can be improved by reducing the set of predictors, beginning with the leave-one-variable-out approach.

Table 4.5: Correlation matrix among macroeconomic factors (EUR). Pearson correlation coefficients ρ (rounded to three decimals). Forward rates and inflation swaps are very highly correlated, and both are positively related to the balance sheet. The balance sheet is strongly negatively correlated with the swap spread, while the curve slope shows negative correlations with forward rates, inflation swaps, and swap spreads.

	FR	**IS**	**BS**	**SS**	**CS**
Forward rate (FR)	1.000	0.805	0.546	-0.134	-0.335
Inflation swap (IS)	0.805	1.000	0.579	-0.160	-0.243
Balance sheet (BS)	0.546	0.579	1.000	-0.521	0.038
Swap spread (SS)	-0.134	-0.160	-0.521	1.000	-0.433
Curve slope (CS)	-0.335	-0.243	0.038	-0.433	1.000

4.2 Leave one feature out (LOFO) importance

As a next step, we assess the marginal importance of each macroeconomic factor by re-estimating the OLS model while omitting one feature at a time. For every expiry–tenor combination, the baseline model is first fitted using all five factors and its mean squared error (MSE) is recorded. Then, for each factor in turn, we refit the regression without that factor and calculate the change in MSE relative to the baseline. A large positive ΔMSE means that the factor is important, as excluding it worsens the fit, while a small or even negative ΔMSE means that the factor adds little explanatory power or may even increase noise.

This procedure gives us a simple way to rank the relative importance of predictors. For example, if removing the balance sheet variable leads to a large increase in MSE, it shows that this factor is crucial to explain volatility. In contrast, if excluding the swap spread reduces the error slightly, it suggests that this factor does not contribute much in that particular expiry–tenor combination. After identifying which features are most important, we then re-fit OLS models using only these factors, with the aim of improving predictive performance.

Figure 4.3 shows the LOFO results for USD swaptions. To provide a clearer overview, we omit the results for 3m and 12m expiries as well as 24m and 84m tenors, since these patterns were very similar to their neighboring grids. For example, the 6m expiry is broadly representative of the results seen at 3m and 12m. At short expiries and tenors, the balance sheet clearly stands out as the only factor whose removal increases the MSE, while the other predictors play little role. Moving towards longer expiries and tenors, the importance becomes more distributed: the curve slope and the inflation swap in particular contribute strongly, while the forward rate and swap spread also show relevance in some regions. This indicates that the drivers of volatility vary across the cube, with the balance sheet dominating at the short end and other factors gaining weight at longer maturities.

For EUR swaptions, shown in Figure 4.4, the pattern differs. As above, we omit the

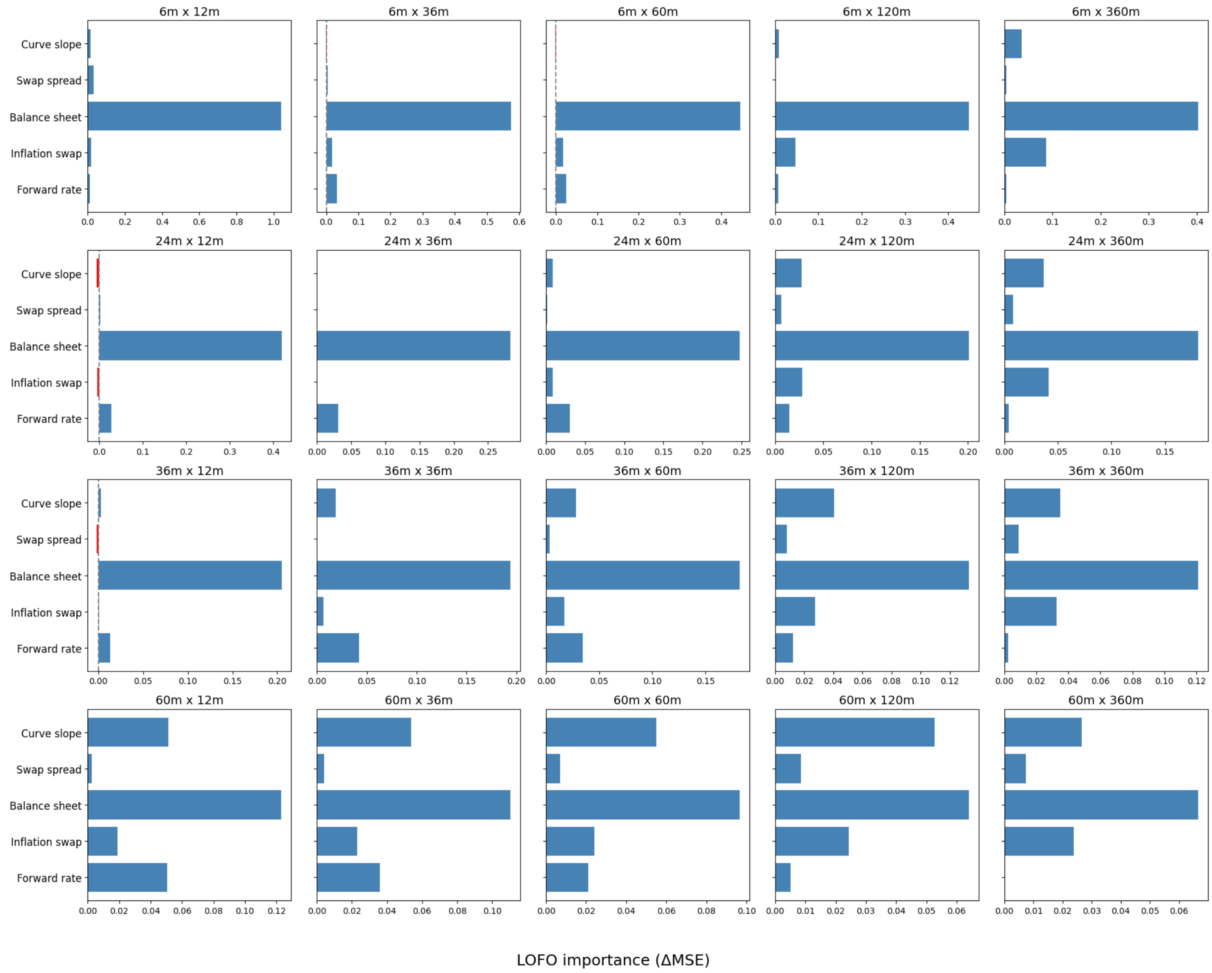

Figure 4.3: Leave one feature out (LOFO) importance for USD swaptions across expiry–tenor combinations. Bars show the increase in mean squared error (MSE) when each factor is removed, with larger values indicating greater importance. The results highlight that balance sheet effects dominate at short expiries and tenors, while at longer maturities the influence becomes more evenly distributed across inflation expectations, curve slope, and forward rates.

3m and 12m expiries as well as the 24m and 84m tenors to improve clarity. Across much of the cube, the forward rate dominates as the most relevant factors, particularly at medium and long tenors. The balance sheet contributes in some regions, but its role is clearly less pronounced than in the USD case. The curve slope, swap spread and inflation swap show importance only in selected combinations, rather than more broadly across the surface. Overall, the EUR results indicate that rate and inflation expectations are the primary drivers of volatility, while balance sheet and slope effects are comparatively limited.

Comparing the two currencies, the main distinction is that USD volatility is more heavily linked to the balance sheet and curve slope, particularly at short and medium maturities, while EUR volatility is more strongly tied to the forward rate and inflation swap. The swap spread remains the weakest factor in both markets, showing only occasional importance in specific regions of the cube.

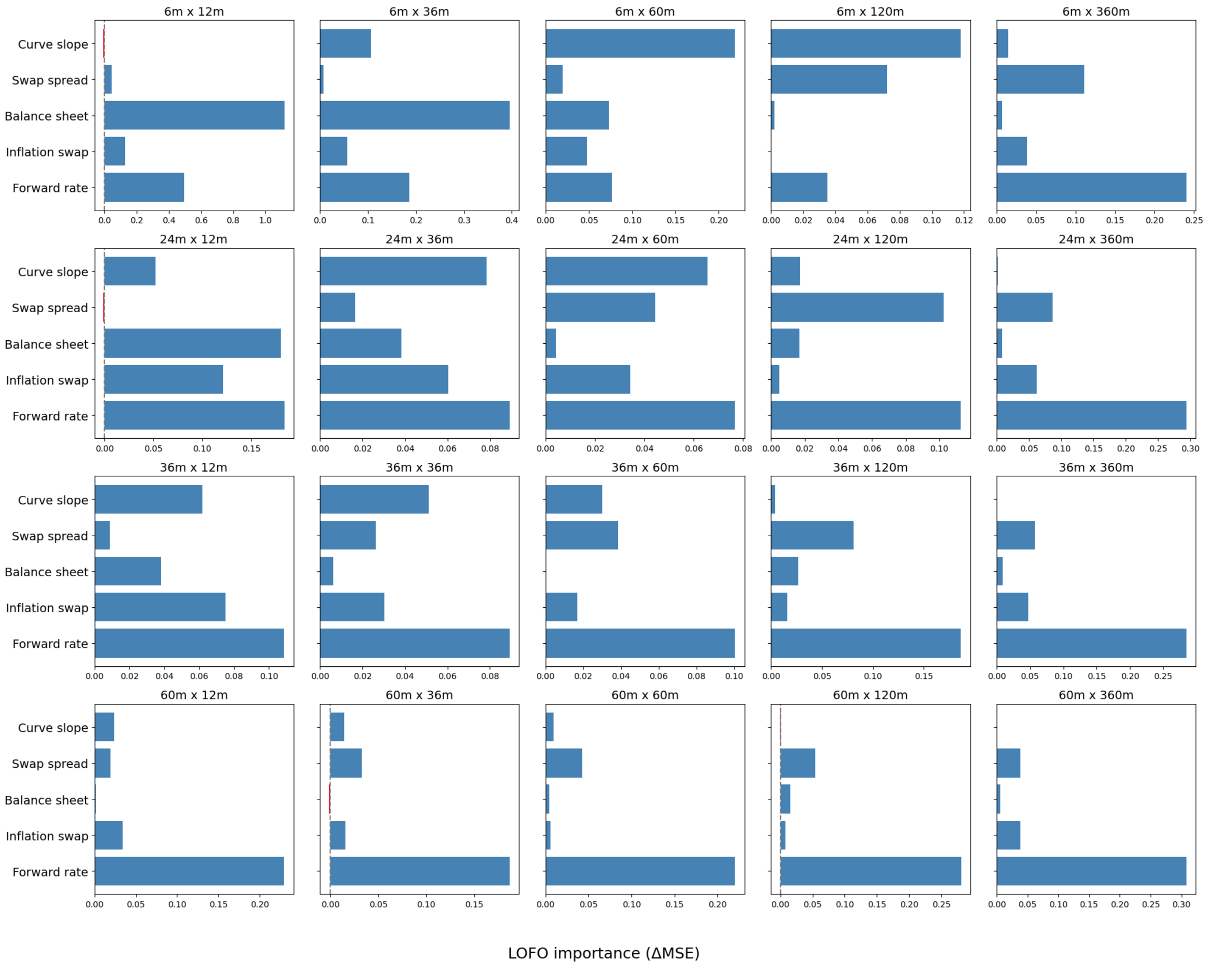

Figure 4.4: Leave one feature out (LOFO) importance for EUR swaptions across expiry–tenor combinations. Bars show the increase in mean squared error (MSE) when each factor is removed, with larger values indicating greater importance. The results indicate that forward rates are the main driver across most of the cube, and they dominate almost exclusively at longer expiries and tenors, while the other factors contribute only in specific regions.

To test whether reducing the number of predictors improves performance, we re-fit OLS models excluding those factors that showed a negative contribution in the LOFO analysis. In other words, if leaving out a factor reduced the mean squared error, we dropped it from the specification, since its inclusion was not helping the fit. To make the comparison fair, we only looked at expiry–tenor combinations where LOFO actually removed predictors and compared the baseline model with all five factors against the reduced model.

For USD, this was the case in eight regressions, where one or two of the five predictors were dropped. LOFO most often removed the the curve slope, while the inflation swap was excluded twice at medium expiry–tenor combinations. The forward rate and the balance sheet were never removed. Across these eight regressions, the mean out-of-sample R^2 increases from 0.4320 to 0.4336, the adjusted R^2 from 0.4264 to 0.4294, and the mean squared error decreases from 0.8693 to 0.8668.

For EUR, LOFO removed variables in eight regressions. The curve slope was excluded most frequently, while the inflation swap was dropped once, the balance sheet twice, and the swap spread once. The forward rate, by contrast, was always retained. Across these these eight regressions, the mean out-of-sample R^2 improves slightly from 0.7549 to 0.7558, the adjusted R^2 from 0.7510 to 0.7528, and the MSE decreases from 0.5879 to 0.5850.

Overall, the LOFO results show that not all factors contribute equally to explaining volatility. It is particularly interesting to see that removing variables which negatively affect the LOFO fit can actually improve the model, even if only modestly. In the next section, we move beyond this manual exclusion and apply Lasso, which selects relevant predictors in a more systematic way.

4.3 Feature selection with Lasso and OLS

As a next step we apply the Lasso method to perform automatic feature selection. Lasso penalizes the size of coefficients and can shrink some of them to exactly zero, effectively removing the corresponding factor from the regression. For each expiry–tenor combination we fit a Lasso model and then re-estimate OLS using only the features that remain non-zero. This allows us to test whether the model can be simplified without losing explanatory power.

For USD swaptions, Lasso sets several coefficients to zero in specific regions of the cube. Table 4.6 lists the combinations where at least one feature was dropped. The factors most often removed are the swap spread, the inflation swap, and occasionally the curve slope. This suggests that in parts of the cube these predictors contribute little and may even add noise. By contrast, the forward rate and balance sheet are almost always retained.

Table 4.6: Expiry–tenor combinations where Lasso dropped at least one feature for USD swaptions. The factors most frequently removed are the swap spread and inflation swap, with the curve slope occasionally excluded as well, while the forward rate and balance sheet are consistently retained.

Start (months)	**Tenor (months)**	**Dropped feature(s)**
3	60	Swap spread
3	84	Swap spread
12	24	Inflation swap
12	36	Inflation swap
12	60	Curve slope
24	36	Inflation swap, curve slope
36	24	Inflation swap, swap spread

For EUR swaptions, the situation is different. Here, Lasso does not set any coefficients to zero, meaning that all five macro factors remain in the model across the entire cube. This indicates that each factor carries at least some explanatory power for EUR volatility.

We next compare the error metrics of the Lasso-selected models to the baseline OLS. To ensure a fair comparison, we restrict attention to the subset of regressions where Lasso actually set one or more coefficients to zero (7 out of 42 in the USD case). For these regressions, we re-fit the baseline OLS with the full set of five predictors and compare its performance to the Lasso model with the reduced set. The results show almost no improvement: for USD, the mean out-of-sample R^2 is 0.4672 compared to 0.4679 for the baseline, the adjusted R^2 increases slightly from 0.4627 to 0.4633, and the mean squared error increases from 0.7204 to 0.7211. For EUR, Lasso did not set any features to zero, meaning the model structure is identical to the baseline OLS and the metrics coincide.

Overall, this indicates that removing predictors through Lasso does not improve predictive accuracy. In the next section, we turn to random forests, a nonlinear model, to test whether nonlinearities provide additional explanatory power.

4.4 Feature selection with random forest and OLS

As a final feature selection approach we turn to random forests, a nonlinear model. A random forest is an ensemble of decision trees, where each tree provides a prediction and the forest aggregates them. One advantage is that random forests naturally produce feature importance measures, which indicate how much each variable contributes to reducing prediction error across the ensemble. The importances are normalized to sum to one, which allows for direct comparison across features.

Figure 4.5 shows the feature importances for USD (blue) and EUR (orange) swaptions across expiry–tenor combinations. Several patterns stand out. For USD, the forward rate and the balance sheet are the two most prominent factors across the cube. At medium expiries and medium-to-long tenors, the inflation swap also becomes relevant, while the curve slope and swap spread remain consistently less important. For EUR, the patterns vary more strongly with maturity. At short expiries, the balance sheet dominates almost entirely, whereas from medium expiries onward the forward rate emerges as the main driver. In addition, for short expiries combined with medium tenors, the curve slope plays a noticeable role, while the swap spread remains negligible throughout.

These results show that nonlinear methods recover a broadly similar factor relevance to OLS and LOFO, but they also highlight structural differences across currencies. USD volatility is tied more closely to balance sheet conditions and, at times, to inflation expectations, whereas EUR volatility exhibits a sharper shift from balance sheet dominance at the short end to forward rate dominance at longer horizons.

We then use random forest importances to select features with an importance above

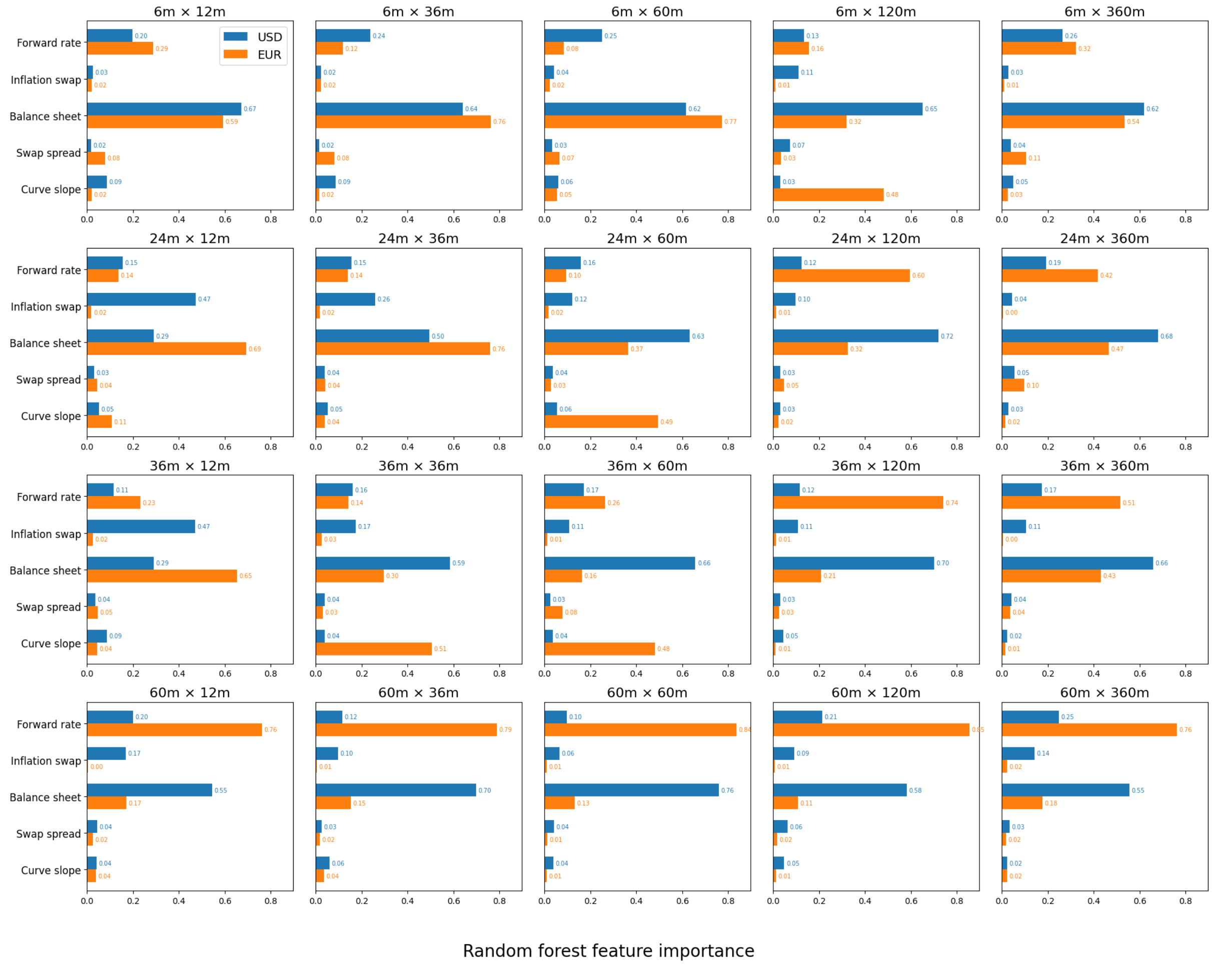

Figure 4.5: Random forest feature importances for USD (blue) and EUR (orange) swaptions across expiry–tenor combinations. Bars show the relative importance of each macro factor, normalized to sum to one per model. The results indicate that USD volatility is driven mainly by balance sheet conditions and occasionally by inflation expectations, while EUR volatility shifts from balance sheet influence at the short end to forward rate dominance at longer maturities.

0.02 and re-fit OLS using only these predictors. For USD, this leads to four regressions where one feature is removed at medium expiries and tenors, twice the inflation swap and twice the swap spread. For EUR, the effect is stronger: in 30 regressions between one and three features are excluded, with more features dropped at longer maturities. In both cases, the forward rate and balance sheet remain consistently included.

We then assess whether this feature selection improves predictive performance. For USD, the mean out-of-sample R^2 decreases slightly from 0.5078 to 0.5014, the adjusted R^2 from 0.5030 to 0.4975, and the mean squared error increases from 1.0886 to 1.1032. For EUR, the mean R^2 falls from 0.6682 to 0.6378, the adjusted R^2 from 0.6642 to 0.6345, and the MSE rises from 0.4967 to 0.5361.

Overall, random forest provides an interesting nonlinear benchmark for feature relevance. However, re-fitting OLS with random forest-selected features does not improve performance. The differences in feature importance across currencies and maturities will be discussed in the next section.

5 Discussion

The results show that the link between macro factors and implied volatility depends on both the currency and the regime. While patterns are fairly stable within each market, clear differences emerge between USD and EUR. The following discussion first compare how the models perform across different time windows, then turns to each factor individually to see how it relates to implied volatility in USD and EUR and relates the statistical results to the economic drivers.

5.1 Regime dependence and time windows

Implied interest rate volatility is shaped by its environment and varies across policy regimes, liquidity conditions, and macroeconomic states. Our sample (2014–2024) spans three distinct environments: a phase of near-zero rates (with negative rates in EUR), a pre-COVID normalization, and the post-2021 inflation shock with rapid rate increases. Fitting one linear model with fixed parameters across such diverse regimes is difficult, and the more piecewise the volatility–rate relationship becomes, the less well a single global fit can capture it.

This is reflected in the OLS results. In the full sample, the mean out-of-sample R^2 is 0.482 for USD swaptions and 0.666 for EUR. Restricting the window to 2021-2024 significantly increases the fit to 0.715 (USD) and 0.884 (EUR). The shorter window is more homogeneous, marked by rapid hikes and elevated volatility, which makes it easier to map macroeconomic factors to volatility. The recent JP Morgan study focuses on this same period for exactly that reason [5]. This is also consistent with recent Federal Reserve evidence that the rise in option-implied volatility during 2021–24 coincided with greater downside risks to activity, which helps explain why macro drivers load more strongly in this regime [4].

Tables 5.1 and 5.2 compare LOFO, Lasso, and random forest across three windows (2014–2024, 2014–2020, 2021–2024). Following the same procedure as in the results section, we compute metrics only for expiry–tenor combinations where the predictor set differs between methods. For these cases, we compare the three feature-selection approaches against a re-estimated OLS baseline with all five factors. Across all windows, feature selection changes R^2 only marginally. LOFO yields small increases in R^2, Lasso leaves the fit essentially unchanged, and random-forest-based selection can even reduce the fit in the longer window. Hence, the variation in R^2 is driven primarily by the chosen time window and its macro regime, not by shrinking the predictor set. While R^2 levels are higher in the two narrower windows, the gap between the baseline OLS and the selected and re-fitted OLS remains similarly small in each window.

Table 5.1: USD: Out-of-sample R^2 (Baseline vs Model) by time period. Results show that feature selection has little effect on explanatory power, with R^2 driven mainly by the time window. Fit improves substantially in the more homogeneous 2021–2024 period.

	2014–2024		2014–2020		2021–2024	
Model	Base	Model	Base	Model	Base	Model
LOFO	0.432	0.434	0.520	0.523	0.754	0.757
Lasso	0.468	0.467	0.723	0.723	0.736	0.737
Random forest	0.508	0.501	0.822	0.626	0.775	0.762

Table 5.2: EUR: Out-of-sample R^2 (Baseline vs Model) by time period. Feature selection again produces only marginal changes, while model fit is strongly dependent on the chosen regime. R^2 is highest in the narrower windows, particularly 2014–2020 and 2021–2024.

	2014–2024		2014–2020		2021–2024	
Model	Base	Model	Base	Model	Base	Model
LOFO	0.755	0.756	0.890	0.892	0.874	0.875
Lasso	0.755	0.756	0.864	0.864	0.863	0.863
Random forest	0.668	0.638	0.878	0.869	0.875	0.867

These results underline that OLS with five macro factors is not well suited to explain volatility across a decade that spans such different regimes. On narrower, more homogeneous windows the regressions perform a lot better.

5.2 Macroeconomic factors and volatility dynamics

This section reads the OLS coefficient heatmaps (Figures 4.1–4.2) together with the simple correlations between factors and residuals (Table 4.3 and the factor–factor matrices in Tables 4.4–4.5). For each factor, we first summarize what the data show across expiry–tenor and currencies, and then discuss why the sign and pattern are plausible from a macroeconomic perspective. We also draw on feature importance measures from the leave one feature out (LOFO) analysis, Lasso, and random forest to complement the OLS results.

5.2.1 Forward rate (rate level / policy stance)

The forward rate shows clear differences across currencies. For USD the OLS heatmap is mostly positive away from the short corner and peaks at intermediate expiry–tenor pairs, while the residual correlation is mildly positive ($\rho \approx 0.112$). The scatter plots, which show the relationship between the volatility and each macroeconomic factor for several expiry–tenor combinations, in Figure 3.4 confirm this pattern: higher forward rates are generally associated with higher implied volatility. For EUR, on the contrary, the coefficients are negative in most of the cube and the correlation is strongly negative ($\rho \approx -0.409$). This regime effect reflects the long periods when EUR rates were at or below the effective lower bound (ELB), the point where policy rates cannot be lowered further without losing effectiveness. In such an environment, higher forward rates often signaled greater policy certainty or a gradual exit from the ELB, which reduced implied volatility [3].

Looking across maturities, the pattern makes sense. At the short end, volatility mainly reflects near-term policy expectations; when these are clear, residual volatility is low and coefficients are small or negative. Medium expiries and tenors are most sensitive because they capture both immediate policy risk and medium-term uncertainty, which explains the peaks in the heatmaps and scatter plots. At the long end, the link weakens again as rates further out are anchored by inflation targets and long-run yields.

The feature-importance results support this view. In USD, forward rates are not the main driver and only matter at some long-expiry points. In EUR, they play a bigger role, especially at long expiries where they are often the most important predictor. This suggests that in the U.S. they are a secondary, maturity-dependent factor, while in the euro area they are central for medium and long-expiry swaptions, reflecting the impact of the ELB and their close link with inflation expectations.

5.2.2 Inflation swap (long-term inflation expectations)

The 5yx5y inflation swap, a market implied measure of expected inflation over a five-year period starting five years ahead, is widely used as a proxy for long-term inflation expectations. In the OLS results, coefficients differ sharply across currencies: they are mostly negative for USD and positive for EUR (Figures 4.1–4.2), with correlations showing the same pattern. Figure 3.4, where the second column plots implied volatility against inflation expectations, indicates that volatility is lowest when expectations are near 2% and rises as they move away from this level.

For the USD, the negative coefficients are somewhat counterintuitive. Higher long-term inflation expectations would normally be expected to raise volatility by increasing uncertainty around the Fed's future rate path. Federal Reserve research highlights inflation uncertainty, which is closely linked to expectations, as a key driver of short-term rate volatility over the past three decades [3]. Implied volatility typically rises when expectations move above target, consistent with the idea that deviations from 2% increase policy uncertainty. Related work also finds that the squared deviation of the inflation swap from 2% has a positive coefficient on volatility, reinforcing this link [6]. By contrast, our regressions suggest that in this sample, higher inflation swaps were associated with lower residual volatility. One explanation is that much of the variation in USD volatility was already absorbed by the forward rate, which is highly correlated with the inflation swap, leaving a negative residual link in the regression.

For EUR, the positive coefficients at short and medium maturities fit the macroeconomic intuition. In 2021–22, elevated inflation expectations coincided with greater uncertainty about the effective lower bound's policy timing[3]. In this setting, higher expectations naturally raised implied volatility, especially after a long period near or below the ELB. Rising expectations signaled a shift away from the low-rate regime and pushed up the volatility risk premium. This interpretation is consistent with recent modeling evidence from BNP: a EUR 10y20y rates volatility model, which estimates fair-value implied volatility for swaptions, includes 10-year inflation swaps as a key driver and finds they consistently explain long-term rates volatility. For short-term EUR volatility, inflation swaps were even identified as the single best explanatory factor [7].

In terms of importance, inflation expectations play a larger role in EUR than in USD. While in USD they appear as a secondary driver, in EUR they explain a greater share of the volatility pattern across expiries and tenors. This difference matches the broader evidence. In the US, inflation uncertainty matters but competes with other drivers like the balance sheet and growth outlook. In EUR, by contrast, inflation swaps have been one of the most consistent explanatory variables for long-term rate volatility.

5.2.3 Balance sheet (QE/QT and liquidity conditions)

The balance sheet variable, measured as the total assets held by the Federal Reserve for USD and by the ECB for EUR, is used here as a proxy for QE stance and overall market liquidity. Regression coefficients are small only because balance sheets are measured in trillions, so even a small value still implies a large effect. In general, larger balance sheets are linked to lower volatility, consistent with the view that QE adds liquidity and dampens rate moves (Figure 3.4). In the heatmaps, coefficients are close to zero in both currencies but tend to be slightly higher at short maturities. This indicates that short-dated swaptions react most to balance sheet changes, especially around new QE/QT announcements. The positive signs are in line with JP Morgan's empirical study [5].

The feature importance results confirm that the Fed's balance sheet has been a structural driver of USD volatility. Its relatively strong negative correlation with residual USD volatility ($\rho \approx -0.515$) suggests that expansions in the Fed's balance sheet were often matched by lower volatility. In EUR, the effect is less consistent: the balance sheet matters only in some short-tenor cases, and the overall correlation is close to zero. This supports the view that European volatility has been shaped more by stress episodes than by the ECB's balance sheet itself.

Overall, the evidence points to the balance sheet as a key driver of USD volatility, while in the euro area its role has been weaker and more episodic.

5.2.4 Swap spread (funding and credit premia)

The 10-year swap spread, defined as the difference between the 10-year swap rate and the 10-year government bond yield, is often used as a proxy for funding costs, credit risk, and market liquidity. In the heatmaps, USD coefficients are mostly positive, which is intuitive: wider swap spreads tend to reflect higher funding premia or balance-sheet stress, both of which increase uncertainty and push up implied volatility. By contrast, EUR coefficients are almost uniformly negative. In EUR, the negative pattern likely reflects market features, where persistent negative swap spreads have often coincided with stable or lower volatility.

Figure 3.4 reinforce this view: volatility is lowest when swap spreads are near zero, implying normal market condition, and increases when spreads move far into negative or positive territory, reflecting market stress and liquidity changes. The correlations tell a similar story: in EUR, the swap spread has a stronger negative link with volatility ($\rho = -0.398$), while in USD the relationship is close to zero. The feature importance results reinforce this finding. In USD, the swap spread has very low importance in both LOFO and random forest results, with little explanatory power compared to other macro drivers. In EUR, importance is also generally low but occasionally rises for short-expiry, long-tenor contracts, suggesting that swap spreads captured elements of stress in certain parts of the cube.

Overall, swap spreads play a minor role relative to forward rates, inflation expectations, or balance sheet policy. They behave intuitively for USD, where wider spreads go hand in hand with higher volatility, but in EUR the relationship is weaker, more negative, and driven by structural features of the euro swap market.

5.2.5 Curve slope (near-term policy path)

The curve slope factor (3mx3m–1yx3m) captures the steepness of the very front end of the forward curve and serves as a proxy for near-term policy expectations. Intuitively, volatility is lowest when the curve is flat, reflecting broad agreement about the central bank's path. Volatility rises as the slope becomes strongly positive or negative, signaling uncertainty around regime shifts or stress. The scatter plots in Figure 3.4 show a clear

U-shape: volatility is lowest when the slope is flat and rises as it becomes strongly positive or negative.

The heatmaps highlight different patterns across currencies. For USD, coefficients are negative at short expiries, meaning that when the front-end steepens, near-term swaption volatility tends to fall. This can be read as a confidence effect: a steeper curve indicates a clearer policy trajectory, reducing immediate uncertainty. At longer maturities the coefficients turn positive. This suggests that steepening is linked to higher long-term volatility, as markets price more macro risk premia. For EUR, coefficients are positive across the cube, especially for short-expiry, medium-tenor contracts, implying that steepening was generally associated with higher implied volatility. This fits the post-ELB regime, where curve steepening often coincided with uncertainty about policy normalization.

The feature importance results show that the slope has little explanatory power in USD, while in EUR it is more relevant, especially at medium and long maturities. This suggests the slope explained more of the volatility in Europe than in the US.

These results are in line with earlier work, which often uses short-end slope factors to capture policy-rate uncertainty, especially at short expiries [6]. In USD, the negative relation at the very front end contrasts with the positive effect at longer maturities, showing the slope can signal both near-term policy clarity and longer-term uncertainty. In EUR, the positive relation across the cube suggests that steepening was generally read as a sign of higher volatility, reflecting policy regime shifts and the greater sensitivity of short- to medium-dated products.

6 Conclusion and future work

6.1 Summary

This thesis asked whether interest rate volatility can be explained by more than just the level of rates. We studied this question using a two-stage framework on USD and EUR swaptions data from 2014 to 2024. In Stage 1, we fitted a volatility–rate backbone to remove the dependence of volatility on rate levels, and in Stage 2, we explained the remaining variation with macroeconomic factors.

We draw three main conclusions from running separate regressions for each expiry–tenor pair, comparing an OLS baseline with LOFO, Lasso, and random forest. First, explanatory power depends heavily on the regime. Models fitted across the full decade perform only moderately well, but they fit much better in 2021 and after, when both rates and volatility were high. This shows that a single global model cannot capture such heterogeneity and that regimes matter. Second, only a few macro drivers consistently explain volatility. Forward rates, inflation expectations, and central bank balance sheets stand out, while swap spreads and the curve slope are only sometimes relevant. Forward rates dominate in EUR, balance sheet policy is most important in USD, and inflation expectations matter more once the euro area moved away from the effective lower bound. Third, moving beyond OLS does not significantly improve model performances. Lasso, LOFO, and random forest confirm the same core drivers without increasing the out-of-sample fit. Their value is more in showing which factors matter for different maturities and in confirming that the same few variables drive results across methods. Overall, what matters more is choosing homogeneous time windows and the right factors rather than adding model complexity.

Overall, the results show that implied volatility depends both on the level of rates and on macro factors that vary by regime. The two-stage setup helps to separate these effects and gives a way to compare sensitivities across maturities. For this thesis, the main takeaway is that volatility cannot be explained by rate levels alone and that changes in the macro environment matter a lot.

6.2 Limitations

The results should be read with several limitations in mind. These mostly reflect the sample, the way the backbone and regressions were designed, and how the models were evaluated. Mentioning these limits also helps to motivate the next section on future work.

Sample period and currency scope. The sample covers 2014 to 2024, which spans low rates, the effective lower bound in EUR, and the post-pandemic inflation shock. This gives variation across regimes, but it excludes earlier high-rate cycles. Therefore, the results may not capture the volatility–rate link in very high-rate environments, so the findings mainly reflect the past decade. The analysis is also limited to USD and EUR. These are the most liquid markets and a good starting point, but they share institutional features. Other markets such as GBP or JPY could behave differently due to policy frameworks, liquidity, or market structure. Generalization beyond USD/EUR should therefore be cautious.

Backbone specification and two-stage structure. In Stage 1, we use a single exponential backbone across all expiries and tenors. This gives a stable and smooth mapping from yields to volatility, but it only approximates the three-regime pattern discussed in the literature. A piecewise fit, rather than a single smooth function, may capture the low–mid–high shape better, especially at the extremes. The two-stage framework also assumes a clear split between the base effect (the backbone) and the macro drivers (Stage 2). In practice, these are not fully separate: forward rates, inflation, and balance sheet policy often move together, and some of this influence may already be absorbed by the backbone. This overlap makes it harder to argue that Stage 2 reflects only macro effects.

Macroeconomic factors and multicollinearity. In Stage 2, we use five macro variables: the forward rate, long-term inflation swap, central bank balance sheet, the 10-year swap spread, and a front-end slope. This choice keeps the model simple and allows us to keep daily frequency, but it leaves out slower-moving fundamentals or liquidity measures that could matter at some horizons. Even within the five variables, some are strongly correlated (for example forward rates and inflation swaps), which makes the coefficient estimates unstable. Because of this, in our analysis we focus more on the sign of coefficients than on their exact size. Still, multicollinearity makes it difficult to interpret individual betas as structural effects.

Window dependence and out-of-sample fit. The explanatory power changes a lot across regimes. In the full 2014-2024 window, the average out-of-sample R^2 is modest (below 0.5 in USD and below 0.7 in EUR), while narrower windows such as 2021–2024 perform much better. This suggests that a single global linear model struggles across different regimes. It also means that our reported fits depend on the chosen evaluation window, and we cannot claim stability over longer horizons based only on the last decade.

Train–test splitting in a time-series setting. For the evaluation we used an 80/20 random split instead of a chronological split. Random splits are common in cross-sectional data, but in time-series with regime shifts they can mix different states and lead to overly optimistic error estimates. A chronological split would have placed most of the high-volatility period (the last two to three years) in the test set, which would have been a tougher and more realistic forecast task. We would have preferred to use a chronological split, but we chose random splits to keep results comparable across many expiry–tenor regressions. This is still a limitation, since the out-of-sample metrics may not reflect true real-time performance. More time-aware approaches, such as blocked or rolling splits, would be better suited for future work.

Overall, these limitations do not change the main results, but they do narrow their scope. In this thesis we worked with USD and EUR data from the past decade, a single exponential backbone, and a small set of daily macro factors. In the next section we discuss what other choices could have been considered if more time and data had been available.

6.3 Future directions

There are several ways in which this framework could be extended, ranging from refining the methodology, a broader macroeconomic set and more flexible models.

Broader macro set. We kept the macro regressions simple by focusing on five liquid, daily variables: forward rates, inflation swaps, central bank balance sheets, swap spreads, and a curve slope. A natural next step would be to add slower-moving fundamentals such as GDP growth, unemployment, or financial conditions indices. These are not available at daily frequency and are often released with a lag, but simple nowcasting or mixed-frequency methods could still make them useful for medium-term forecasts.

Currency coverage. We worked with USD and EUR because they are the deepest and most important swaption markets. Testing the same framework in GBP or JPY would be useful, since these markets face different policy settings (for example, yield curve control in JPY). This would show whether the results hold more broadly or are specific to USD and EUR.

Backbone and model extensions. In Stage 1 we used a single exponential backbone. This gave a stable fit, but a piecewise or regime-switching backbone could capture the low–mid–high shape discussed in the literature more accurately, especially at the extremes. In Stage 2 we worked with OLS, LOFO, Lasso, and random forests. A natural next step would be to try other methods such as gradient boosting, support vector regression, or neural networks. These can capture stronger non-linearities, but the trade-off is that they are harder to interpret. The question then becomes whether this additional complexity really improves out-of-sample accuracy compared to the simpler models.

Evaluation design. Finally, we would change how the models are tested. In this thesis, we used random 80/20 splits, but with more time we would prefer rolling splits that follow the time order. This would give a clearer picture of how the models perform when regimes shift and would be closer to how they would be used in practice.

In short, future work could look at a wider set of factors, test other currencies, try different backbone fits, and use time-aware evaluation. This would build on the framework in this thesis and show how far the findings apply in other settings.

Bibliography

[1] Nick Deguillaume, Riccardo Rebonato, and Andrey Pogudin. The nature of the dependence of the magnitude of rate moves on the rates levels: a universal relationship. *Quantitative Finance*, 13(3):351–367, 2013. doi:10.1080/14697688.2012.740569.

[2] Riccardo Rebonato and Amir Aouadi. How do the volatilities of rates depend on their level? The "universal relationship" revisited. *Journal of Fixed Income*, 30: 17–30, 2021. doi:10.3905/jfi.2021.1.110.

[3] Cisil Sarisoy. Drivers of option-implied interest rate volatility. FEDS Notes, Board of Governors of the Federal Reserve System, October 2024. URL `https://www.federalreserve.gov/econres/notes/feds-notes/drivers-of-option-implied-interest-rate-volatility-20241024.html`.

[4] Cisil Sarisoy. Elevated option-implied interest rate volatility and downside risks to economic activity. FEDS Notes, Board of Governors of the Federal Reserve System, December 2023. URL `https://www.federalreserve.gov/econres/notes/feds-notes/elevated-option-implied-pinterest-rate-volatility-and-downside-risks-to-economic-activity-20231222.html`.

[5] J.P. Morgan Global Rates Strategy Team. Empirical volatility model: Vol-rate empirical model with external factors. Interest rate derivatives 2025 outlook: "nobody said it was easy", J.P. Morgan, 2024. JP Morgan internal research note.

[6] JP Morgan Global Rates Strategy Team. Vol-versus-rate directionality. Technical report, JP Morgan, 2022. JP Morgan internal research note.

[7] Ruoxi Sun and Michael Sneyd. Ev-vol series: Introducing eur 10y20y rates volatility model. Technical report, BNP Paribas Global Markets, 2025.

[8] Mark Kritzman. What practitioners need to know about estimating volatility part 1. *Financial Analysts Journal*, 47(4):22–25, 1991. doi:10.2469/faj.v47.n4.22.

[9] Jaehyuk Choi, Minsuk Kwak, Chyng Wen Tee, and Yumeng Wang. A Black-Scholes user's guide to the Bachelier model. *Journal of Futures Markets*, 42(5):959–980, 2022. doi:10.1002/fut.22315.

[10] Peter Jäckel. By implication. *Wilmott Magazine*, pages 60–66, 2006.

[11] Peter Jäckel. Let's be rational. *Wilmott Magazine*, pages 40–53, 2015. doi:10.1002/wilm.10395.

[12] Yiran Cui, Sebastian del Baño Rollin, and Guido Germano. Stability of calibration procedures: fractals in the Black-Scholes model, 2016. URL `https://arxiv.org/abs/1612.01951`.

[13] Yiran Cui, Sebastian del Baño Rollin, and Guido Germano. Comp0043 numerical methods for finance – Lecture 9: Model calibration, 2020. Lecture slides.

[14] Fischer Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3(1-2):167–179, 1976. doi:10.1016/0304-405X(76)90024-6.

[15] Patrick S. Hagan, Deep Kumar, Andrew S. Lesniewski, and Diana E. Woodward. Managing smile risk. *Wilmott Magazine*, 1:84–108, 2002.

[16] Riccardo Rebonato. *Modern Pricing of Interest-Rate Derivatives: The LIBOR Market Model and Beyond*. Princeton University Press, 2002. ISBN 978-0-691-09049-8. doi:10.2307/j.ctt7rpkk.

[17] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973. doi:10.1086/260062.

[18] Jaehyuk Choi, Kwangmoon Kim, and Minsuk Kwak. Numerical approximation of the implied volatility under arithmetic Brownian motion. *Applied Mathematical Finance*, 16(3):261–268, 2009. doi:10.1080/13504860802583436.

[19] Jim Gatheral. *The volatility surface: a practitioner's guide*. Wiley Finance. John Wiley & Sons, Hoboken, NJ, 2006. ISBN 978-0-470-02511-9.

[20] Stewart Mayhew. Implied volatility. *Financial Analysts Journal*, 51(4):8–20, 1995. doi:10.2469/faj.v51.n4.1916.

[21] Bruno Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[22] Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2): 327–343, 1993. doi:10.1093/rfs/6.2.327.

[23] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer, New York, 2013. ISBN 978-1-4614-7137-0. doi:10.1007/978-1-4614-7138-7.

[24] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition, 2009. ISBN 978-0387848570. doi:10.1007/978-0-387-84858-7.

[25] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. doi:10.1111/j.2517-6161.1996.tb02080.x.

[26] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall/CRC (Taylor & Francis Group), 1st edition, 2015. ISBN 9781498712163.

[27] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi:10.1023/A:1010933404324.

[28] Christoph Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Self-published, 3 edition, 2025. URL `https://christophm.github.io/interpretable-ml-book/`.

[29] Christoph Molnar. Leave one feature out (LOFO) importance. `https://christophm.github.io/interpretable-ml-book/lofo.html`, 2025. Online chapter from Interpretable Machine Learning.

[30] Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018. doi:10.1080/01621459.2017.1307116.

[31] Karl Pearson. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58:240–242, 1895. doi:10.1098/rspl.1895.0041.