Interfaces with Other Disciplines

Reducing transaction costs for interest rate risk hedging with stochastic programming

Jörgen Blomvall*, Johan Hagenbjörk

Department of Management and Industrial Engineering, Linköping University, Linköping SE-58183, Sweden

A R T I C L E   I N F O

Article history:
Received 29 August 2020
Accepted 1 February 2022
Available online 5 February 2022

MSc:
91G20
91G30
90C19

Keywords:
Finance
OR in banking
Risk management
Stochastic programming

A B S T R A C T

Traditional methods for hedging interest rate risk do not take transaction costs into account as they aim to eliminate all risk. We propose a two-stage stochastic programming model for hedging interest rate risk where transaction costs are weighed against portfolio variance. High-quality measurements of term structures enable us to extract the systematic risk factors and make precise estimates of the perceived transaction costs. The hedging cost is weighed against the reduction in portfolio variance by using an adjustable hedging parameter. The hedging procedure is simulated on a daily basis in a realistic setting over an out-of-sample period from 2002 to 2018, and the results are compared to traditional hedging methods through detailed performance attribution. Using second-order stochastic dominance, we show that the proposed method is preferred by all risk-averse investors.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

In the second half of 2018, interest rate swaps (IRS) accounted for $327 trillion out of the total outstanding amount of $427 trillion in the global interest rate derivative market (Bank of International Settlements, 2019). These instruments are traded on the OTC market, where major banks act as market makers. As a result of trades with customers, banks hold trading portfolios of these interest rate derivatives. Banks earn money from having their customers cross the bid-ask spreads when entering a contract. However, market makers usually have no desire to be exposed to interest rate risk, which is why banks usually hedge their positions by trading with other counterparties. With the upcoming fundamental review of the trading book (FRTB) regulation by Basel Committee on Banking Supervision (2016), capital requirement for risk exposures in these trading books is increased, providing an even stronger incentive to hedge the risk.

Hagan & West (2006) describe a method to delta hedge interest rate risk through “Bumping”, “Boxes” and “Waves”. This type of delta hedging determines the hedge by finding the portfolio where the first-order partial derivatives with respect to each risk factor are equal to zero. However, such delta-hedge strategies neither consider transaction costs nor the impact size of different systemic risk factors. While trading swaps with another counterparty is not associated with direct transaction costs, it involves crossing the bid-ask spread. This implies paying (receiving) higher (lower) interest rates than what would be considered a fair market yield. Conducting a trade generally results in an initial positive net present value (NPV) for the market maker and an equally large negative NPV for the customer. We define the perceived cost of entering a trade as the price difference given the quoted yield and the fair market yield, which is an unobservable value that must be measured using a theoretical pricing model. Thereby, the fair yield will depend on the term structure measurement method used. An exact method based on bootstrapping mid yields will reprice all in-sample instruments exactly, implying that the fair market yield coincides with the midpoint of the quotes. Inexact methods, on the other hand, allow prices to deviate from the observed market quotes. This deviation creates an asymmetry between the perceived costs of entering a certain contract in its paying or receiving direction. Given an accurate term structure measurement method, this asymmetry can be used to estimate the perceived transaction costs associated with trading each instrument.

In this paper, we propose a two-stage stochastic programming (SP) model for hedging interest rate risk while accounting for transaction costs. We use the method developed by Blomvall (2017) to provide estimates of the perceived transaction costs, which are weighed against portfolio variance using an adjustable hedging parameter. The hedging is carried out on a daily basis in a realistic simulation setting over the out-of-sample period.
2002–2018. We compare the results of our SP model to a naïve method and the traditional hedging methods developed by Hagan & West (2006). By utilizing the residual-free performance attribution framework by Blomvall & Hagenbjörk (2019), we can provide an accurate and detailed ex-post explanation that shows that the improved hedging performance is the result of reduced perceived transaction costs. We study the distribution of daily earnings after applying the hedging methods to a market maker’s swap book. The first two moments, average daily earnings and portfolio volatility, are studied using pairwise statistical tests. Traditional hedging methods produce expensive hedges, resulting in lower average daily earnings than SP models with similar portfolio volatilities. Through the decrease in hedging cost, our proposed method shifts the distribution of daily earnings upwards to such an extent that it dominates the traditional methods by second-order stochastic dominance.

Several SP fixed income applications can be found in the literature, but to the best of our knowledge, there is no hedging application. Similar studies have been carried out in the options market by Gonzio, Kouwenberg, & Vorst (2003) and Barkhagen & Blomvall (2016), where the SP models have outperformed traditional ones. Wu & Sen (2000) present a similar study on currency hedging where they find that their SP model has significant advantages over traditional models. Davari-Ardakani, Aminnayeri, & Seifi (2015) study a multi-stage SP model to hedge a stock portfolio using stock options and achieve an improved efficient frontier. On the fixed-income side, the Black, Derman, & Toy (1990) model have been used to model interest rate stochasticity by Worzel, Vassiadou-Zeniou, & Zenios (1994), Zenios (1995), Golub, Holmer, McKendall, Pohlman, & Zenios (1995), Vassiadou-Zeniou & Zenios (1996), Zenios, Holmer, McKendall, & Vassiadou-Zeniou (1998), Consiglio & Zenios (2001), to solve various portfolio management problems. This paper thus bridges a gap in the literature by using stochastic programming for fixed income hedging.

The rest of the paper is organized as follows. In Section 2, the hedging problem is addressed by studying the properties of traditional hedging models. In Section 3, the SP model is set up, and Section 4 describes the scenario-generation method. The backtesting simulation and the data used is described in Sections 5 and 6. The results are presented in Section 7 and the paper ends with a discussion and conclusions in Sections 8 and 9.

2. Addressing the hedging problem

This section aims to introduce the hedging problem and the settings in which we approach it. To test our SP model, we simulate trades starting with an empty USD swap book in 2002 and performing daily market-making and hedging to 2018. Such a long backtest is essential since the problem contains a feedback loop where the entered hedge positions each day will affect the portfolio for many years ahead. Each day, we let customers trade to the observed bid and ask closing yields, crossing the spreads. It is assumed that the same yields can be used to take hedging positions with another counterparty, thereby forcing us to cross the same spread. Using a Poisson process that expects a total of 10 trades each day, customers arrive to trade vanilla swaps within the 1–10-year maturity range. The number of trades for each maturity is randomized using the Poisson distribution for both the paying and receiving direction with equal probability. For each trade, the logarithm of the nominal amount is determined by a standard normal-distributed random number, scaled by the base amount of $100.0.1

3 This produces a realistic and challenging setting where customers of varying sizes arrive to trade each day. We use a fixed seed for our random number generator to make the simulation replicable between runs.2

The valuation of the swaps in the portfolio is carried out in a single-curve setting by viewing a swap as a long position in a fixed rate bond, and a short position in a floating rate bond (see Hull, 2006). Given a term structure of interest rates, it is thus easy to compute the NPV by discounting each cash flow. However, term structures are not observable in the market but must be measured through an inverse optimization problem. We use the framework described by Blomvall (2017) to obtain high-quality term structures. The framework is based on a trade-off between smoothness and in-sample price errors through daily discretization and regularization of the term structures in terms of forward rates, as these are best suited for imposing conditions for economic realism. Forward rate innovations observed over the in-sample period 1996–2001 are used for principal component analysis (PCA) to extract the factor loadings (eigenvectors) of the term structure movements (see Litterman & Scheinkman, 1991). The factor loadings are held constant over the entire out-of-sample period to stabilize the hedging decisions and ensuring the same risk-factor interpretation in the performance attribution. By considering only six factors, we can explain almost all variance in spot rates as well as forward rates. These constant factors also generalize well to the out-of-sample period, see (Blomvall & Hagenbjörk, 2019) for further details.

The portfolio value evolution of the unhedged IRS book in Fig. 1 can be analyzed with the residual-free performance attribution developed by Blomvall & Hagenbjörk (2019), see also Table 1. Customer trades, ΔStax, provide $4544 in earnings throughout the period. Since there is no hedging, the associated costs, ΔStax, are zero. The carry term reflects the portfolio’s sensitivity to the passage of time, including both accrued interest and rolloff effects.3 With the current seed, carry reduces the portfolio value considerably during 2009–2011, giving a total loss of $2130. The random changes in the interest-rate curve drive the risk in the portfolio, and for the current seed, the total loss due to risk is $5543. The first six systematic risk factors from the PCA (shift, twist, butterfly, and principal components 4–6) explain almost all the changes in portfolio value through the first-order terms (lower-right panel in Fig. 1). The squared shift term and other second-order terms are all negligible. The performance attribution is based on a second-order Taylor approximation where the lower-left panel in Fig. 1 shows that the effect of higher-order terms, e2, is small. Furthermore, using PCA to reduce the 3650 risk factors to only six has a small effect, the performance attribution of the remaining $3644 insignificant risk factors, e2, is also small. The performance attribution hence allows the total portfolio value of $-3148 to be decomposed into market-making profits ($4544), carry ($-2130), unhedged risk ($-5543), insignificant risk factors ($-4) and errors introduced by the Taylor approximation ($-14). It also provides more detailed information about the impact from the individual risk factors, the evolution of portfolio value over time as well as detailed information about each individual trade.

The approach described in this section sets the scene for the setting in which we study the hedging problem. In the next sec-

---

1 While kept out to maintain lower condition numbers, this figure can be thought of as millions to roughly match the swap book size of a commercial bank.

2 The seed determines when the market maker pays/receives fixed rate, and hence also the earnings from the realized interest rate movements. For pedagogical purposes, we have chosen a seed where the unhedged portfolio experiences a significant loss. The impact on a hedged portfolio is limited since the gains from customer trading remain (these are proportional to the trading volume), while the systematic risks are significantly reduced making the portfolio value growth independent of the customers positions.

3 It is computed analytically from the difference of portfolio values using the same interest rate curve for the current day and by reapplying it the next day.
Fig. 1. Performance attribution of the unhedged portfolio. The portfolio value is the sum of all factors counted from the zero level. The lower-left panel shows the contribution of the error terms in greater detail, and the lower-right panel shows the contribution of each first-order risk factor term, the second-order shift component, and the sum of the remaining second-order risk factor terms.

Table 1
Portfolio value, hedge cost and day-to-day volatility, followed by a specification of the performance attribution components, and a more detailed specification of the risk factors.

<table>
<thead>
<tr>
<th></th>
<th>No Hedge</th>
<th>Naïve Hedge</th>
<th>Bumping</th>
<th>Boxes</th>
<th>Waves</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPV</td>
<td>−3148</td>
<td>951.5</td>
<td>912.0</td>
<td>910.3</td>
<td>926.3</td>
</tr>
<tr>
<td>Daily Earning</td>
<td>−0.7556</td>
<td>0.2484</td>
<td>0.2189</td>
<td>0.2185</td>
<td>0.2224</td>
</tr>
<tr>
<td>Hedge Cost</td>
<td>0</td>
<td>3682</td>
<td>3683</td>
<td>3683</td>
<td>3691</td>
</tr>
<tr>
<td>Volatility</td>
<td>87.97</td>
<td>0.2485</td>
<td>0.01947</td>
<td>0.02011</td>
<td>0.2285</td>
</tr>
<tr>
<td>Carry</td>
<td>−2130</td>
<td>83.09</td>
<td>50.96</td>
<td>49.49</td>
<td>92.9</td>
</tr>
<tr>
<td>ε₁</td>
<td>−3.76</td>
<td>−0.01289</td>
<td>−0.00542</td>
<td>−0.02359</td>
<td>−0.5918</td>
</tr>
<tr>
<td>ε₁²</td>
<td>−13.72</td>
<td>0.01527</td>
<td>0.007071</td>
<td>0.01013</td>
<td>0.08775</td>
</tr>
<tr>
<td>Shift</td>
<td>−2239</td>
<td>9.246</td>
<td>−0.1506</td>
<td>−0.3456</td>
<td>0.9739</td>
</tr>
<tr>
<td>Twist</td>
<td>−1470</td>
<td>−1.471</td>
<td>−0.0653</td>
<td>−0.01356</td>
<td>−0.8084</td>
</tr>
<tr>
<td>Butterfly</td>
<td>−848.3</td>
<td>−1.932</td>
<td>0.01665</td>
<td>0.1806</td>
<td>−0.4263</td>
</tr>
<tr>
<td>4th</td>
<td>−277.2</td>
<td>−0.1591</td>
<td>−0.01311</td>
<td>0.06097</td>
<td>−2.229</td>
</tr>
<tr>
<td>5th</td>
<td>−398.1</td>
<td>−0.3031</td>
<td>−0.112</td>
<td>−0.04475</td>
<td>−12.94</td>
</tr>
<tr>
<td>6th</td>
<td>−2649.6</td>
<td>−0.0295</td>
<td>−0.04663</td>
<td>−0.1041</td>
<td>−3.541</td>
</tr>
<tr>
<td>Shift²</td>
<td>−0.08321</td>
<td>0.0003732</td>
<td>−7.891e-06</td>
<td>−7.927e-06</td>
<td>−3.746e-05</td>
</tr>
<tr>
<td>Σ²nd-rem</td>
<td>−0.001991</td>
<td>0.0001141</td>
<td>−2.428e-06</td>
<td>−2.393e-06</td>
<td>2.072e-05</td>
</tr>
</tbody>
</table>

2.1. Naïve and traditional hedging methods

When introducing bid-ask spreads to the hedging problem, it becomes difficult to efficiently hedge the portfolio risk. A naïve hedging strategy, sometimes referred to as “back-to-back”, for an IRS implies entering a second offsetting IRS with another counterparty. To make it more efficient we study the net effect of the customer positions for each maturity. Traditional methods for hedging interest rate risk include the three methods Bumping, Boxes, and Waves, by Hagan & West (2006, 2008). These methods originate from a setting without spreads, where the aim is to eliminate all risk by solving a system of linear equations. The equations stem
from the risk factor sensitivity of the unhedged portfolio and all the tradable instruments. The methods rely on computing sensitivity measures to different perturbations of the term structure. Bumping implies displacing each quote of the hedging instruments with a small amount, usually one basis point. The term structure is recalibrated using the bumped yield, and the sensitivity is computed as a numerical derivative within what is supposed to be a linear regime of the pricing function. This strategy aligns well with the theory of bootstrapping methods for term structure measurement. Boxes and Waves are two methods where the term structure of forward rates is modified directly using shapes that resemble boxes and triangular waves. Assuming that instruments of increasing maturities \( T_i \), \( i = 1, \ldots, n \) are available for hedging, we create the set \( \mathcal{T} = \{ T_i \}_{i=1}^n \). The Boxes method implies a parallel shift of the forward rate term structure by height \( b \) according to

\[
g_{\text{Boxes}}^{(i)}(t) = \begin{cases} 
0 & t < T_{i-1} \\
 b & T_{i-1} \leq t \leq T_i \\
0 & t > T_i.
\end{cases}
\]  

The name Boxes stems from the box-shaped shift in the term structure. The Waves method uses the same technique where the perturbations form waves with their peaks at the maturity points, \( T_i \in \mathcal{T} \), according to

\[
g_{\text{Waves}}^{(i)}(t) = \begin{cases} 
0 & t \leq T_{i-1} \\
 b \frac{t - T_{i-1}}{T_i - T_{i-1}} & T_{i-1} \leq t \leq T_i \\
 b \frac{T_i - t}{T_i - T_{i-1}} & T_i \leq t \leq T_{i+1} \\
0 & t > T_{i+1}.
\end{cases}
\]  

The methods described by Hagan & West (2008) can be viewed as a first-order Taylor approximation of the portfolio value with respect to the risk factor perturbations. Let \( \Delta \xi \in \mathbb{R}^{n \times 1} \) denote the vector of risk factors innovation and \( P(\Delta \xi) \in \mathbb{R} \) the initial portfolio value. The gradient, \( \mathbf{g}_0 = \nabla_{\Delta \xi} P|_{\Delta \xi=0} \), gives the sensitivities with respect to the risk factors that should be hedged. To hedge the risks, \( n \) instruments are used with prices \( \mathbf{P}(\Delta \xi) \in \mathbb{R}^{n \times 1} \), with the sensitivities in the Jacobian \( \mathbf{G} = \nabla_{\Delta \xi} \mathbf{P}|_{\Delta \xi=0} \). The investment \( \mathbf{x} \in \mathbb{R}^{n \times 1} \) with the total portfolio value \( \mathbf{V}(\mathbf{x}, \Delta \xi) = \mathbf{P}(\Delta \xi) + P(\Delta \xi)^{\top} \mathbf{x} \) gives the first-order Taylor approximation

\[
\Delta \mathbf{V}(\mathbf{x}, \Delta \xi) \approx \Delta \xi^\top \mathbf{G}_0 + \Delta \xi^\top \mathbf{G} \mathbf{x}.
\]  

For the special case when \( m = n = \text{rank}(\mathbf{G}) \), which holds for the three Hagan & West (2008) methods, \( \mathbf{G} \) is a square matrix which forms a linear system where the exposure to the risk factors can be eliminated by setting \( \mathbf{g}_0 + \mathbf{G} \mathbf{x} = 0 \iff \mathbf{x} = -\mathbf{G}^{-1} \mathbf{g}_0 \). By transforming the forward rate risk factors of Boxes and Waves, from (1) and (2), into spot rate risk factors, it is possible to compute analytical sensitivities to the risk factors by differentiating the pricing functions, as demonstrated by Blomvall & Hagenbjörk (2019). We set up the Bumping method using bootstrapping and cubic natural splines in the QuantLib framework (Ametrano & Ballabio, 2000–2019), where each quote is bumped by one basis point when computing the risk factors. Each historical day for the out-of-sample period, (4) determines the rebalanced hedging contracts \( \mathbf{x} \).

The results for the naïve and traditional methods can be found in Table 1, together with a detailed performance attribution of the accumulated results. All portfolios are valued using term structures measured by the Blomvall (2017) method. We see that the total transaction costs are similar for these hedging methods, while the portfolio volatilities differ a lot. The results confirm that Boxes is preferred over Waves, as pointed out by Hagan & West (2008). For this long-term simulation, the results indicate that Bumping is at least as good as Boxes. The naïve back-to-back hedge attains an order of magnitude greater volatility than that provided by the Bumping and Boxes methods. All methods receive some interest for their risk exposures found in the carry term. This term is larger than the realized returns from risk factor movements for all hedging methods but still small compared to the hedge cost. The performance attribution plot for the Bumping method is presented in Fig. 2. We see that the portfolio value increases persistently as a result of customer trades while the risk is close to eliminated, as seen in the lower-right panel. The volatility of the unhedged portfolio, 88, has been decreased to 0.02 (Table 1), which means that essentially all risk has been eliminated by Bumping. However, the hedging costs are substantial.

We propose a two-stage stochastic programming model to provide a more cost-efficient hedging strategy based on a more accurate mathematical modeling of the decision problem. This model is described over the next sections, but first, we give a glimpse of the results when the SP model is used with different risk aversion parameters. We compute excess portfolio value for the SP models over Bumping when different hedging parameters are used. The results are presented in the left panel of Fig. 3, where a logarithmic y-scale is used to enable comparison between methods of different risk aversion. As shown later, the model with hedging parameter \( \alpha = 10^{-4} \) attains a slightly lower variance than the Bumping method, with a steadily increasing difference in portfolio value. An underwater plot for the relative drawdowns of \( \alpha = 10^{-2} \) and Bumping is shown in Fig. 3. We see that the maximum drawdown over the entire period is slightly above \( \$1 \), which can be compared to the average daily earning of \$0.2189 for the Bumping method. These results show that a more realistic optimization model can significantly improve hedging decisions compared to delta hedging. Developing a hedging method with such theoretical performance requires knowledge of the systematic risk factors, their importance for the portfolio risk, and the cost of eliminating each risk factor. These topics will be addressed in the upcoming Sections 3 and 4.

3. Formulating the hedging problem

Both Markowitz (1952) Mean-Variance model and a second-order Taylor approximation of the power utility function give a weighted sum of variance and expected return. Hedging a derivative book differs from traditional asset management as the book value can be negative and the main objective is to reduce risk. With guidance from previous theory, a linear combination of variance and expected return is a reasonable starting point for the objective function. We follow Barkhagen & Blomvall (2016) and replace the expected return with (non-negative) transaction costs to prevent the optimization model from increasing the risk by taking active positions in possibly mispriced instruments. The following proposition proves that the suggested modification of the Mean-Variance model implies that the risk can only be decreased, which is appropriate for hedging applications.

\[4\] The volatility of a portfolio is measured based on the standard deviation of the day-to-day changes in portfolio value. First, we perform customer trades and take hedging positions, both to the closing prices at time \( t \). We then move forward to \( t+1 \) and observe the effects of the movements in the term structure based on this day's closing prices. We hence subtract the transaction costs arising from customer trades, \( \Delta \mathbf{p}^{\text{out}} \), and hedging, \( \Delta \mathbf{t}^{\text{out}} \), when computing volatility of \( \Delta \mathbf{p}^{\text{in}} = \mathbf{R} - \mathbf{R}_{-1} - \Delta \mathbf{p}^{\text{out}} - \Delta \mathbf{t}^{\text{out}} \).

\[5\] Alternative mathematical formulations use either variance or expected return as a constraint. The drawback is that customer demand is exogenous, which combined with the inflexible constraint create a model which do not dynamically balance risk and expected return.
Fig. 2. Performance attribution for the Hagan & West (2006) Bumping method.

Fig. 3. Left panel: Excess portfolio value over Bumping for different hedging parameters determining how much weight needs to be given the transaction costs compared to variance. Right panel: underwater plot displaying the relative drawdown between $10^{-2}$ and Bumping.

**Proposition 1.** For the Mean-Variance model

$$\hat{x} = \arg\min_{x \geq 0} \frac{1}{2} \text{Var}(\bar{W}_{t+1}(x)) + c^T x$$  \hspace{1cm} (5)

it holds that $\text{Var}(\bar{W}_{t+1}(\hat{x})) \leq \text{Var}(\bar{W}_{t+1}(0))$ if $c \geq 0$, when the wealth $\bar{W}_{t+1}(x)$ at time $t+1$ is determined by the decision vector $x$ and the initial holdings are represented by $x = 0$.

Proof.

$$\frac{1}{2} \text{Var}(\bar{W}_{t+1}(\hat{x})) \leq \frac{1}{2} \text{Var}(\bar{W}_{t+1}(0)) + c^T \hat{x}$$

$$= \min_{x \geq 0} \frac{1}{2} \text{Var}(\bar{W}_{t+1}(x)) + c^T x \leq \frac{1}{2} \text{Var}(\bar{W}_{t+1}(0)).$$

The weight, $\alpha$, determines how much emphasis should be put on reducing transaction costs. Each day, the market maker can decide the notional value to enter on the paying side, $x_p \in \mathbb{R}^{N \times 1}$, and receiving side, $x_r \in \mathbb{R}^{N \times 1}$, for the $N$ tradable contracts. Using $\text{Var}(\bar{W}_{t+1})$ to denote the variance of the total portfolio wealth in the next period, $\bar{W}_{t+1}$, we can formulate the optimization model as

$$\min_{x_p, x_r} \frac{1}{2} \text{Var}(\bar{W}_{t+1}) + \alpha (c_p^T x_p + c_r^T x_r)$$

subject to $x_p, x_r \geq 0$.

where $c_p$ and $c_r$ denote the perceived transaction costs for entering the contracts, further described in Section 3.3. The prob-
lem has non-negativity constraints as the bid-ask spread must be crossed when trading.

3.1. The stochastic programming model

Using stochastic programming, we can solve a sample average approximation of (6), where we derive a closed-form expression for \( \text{Var}(W_{i+1}) \), under \( S \) equally probable scenarios. We get the two-stage problem without recourse decisions

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{S} \sum_{i=1}^{S} \left( W_i - \frac{1}{S} \sum_{j=1}^{S} W_j \right)^2 + \alpha (c^T x + c^T x_b) \\
\text{subject to} & \quad W_i = h_i + V^T \beta x_p + V^T \beta x_h, \quad i = 1, \ldots, S \\
& \quad x_p, x_h \geq 0.
\end{align*}
\]

(7)

where \( h_i \) is the value of the current swap book before hedging, \( V_{β} \) and \( V_{β}^T \) are the NPV’s of the available contracts paying or receiving fixed rate and \( c^T \) and \( c^T \) are the contract’s perceived transaction costs. This single-period stochastic programming model has to be solved for each historic day, which makes it important to be able to solve the problem efficiently.

3.2. Solving the stochastic programming problem

For all \( S \) simulated term structure scenarios, we price the portfolio and each of the \( N \) tradable instruments in both the paying and receiving direction. We denote the vector of values for the portfolio holdings for each scenario by \( b \in \mathbb{R}^{S \times 1} \) and create the vector \( x = [x_h, x_p] \in \mathbb{R}^{2N \times 1} \) for our trading decision. We let \( \Sigma \in \mathbb{R}^{2S \times 2S} \) denote the matrix of scenario values for these instruments. The value of the hedged portfolio under the scenarios \( \Sigma \), can be expressed as \( W = \Sigma x + h \). Using matrix notation, the portfolio values subtracted by the average value, present in the objective function in (7), can be expressed as

\[
\tilde{W}_{i+1} = W - \frac{1}{S} \tilde{\Sigma}^T W + \tilde{\Sigma} = \tilde{\Sigma} \tilde{x} + \tilde{h}.
\]

(8)

Given the diagonal probability matrix, \( P \in \mathbb{R}^{S \times S} \), containing the probabilities of each scenario in its diagonal, we can determine the portfolio variance as

\[
\text{Var}(W_{i+1}) = \tilde{\Sigma} \tilde{x}^T P \tilde{x} + 2 \tilde{h} \tilde{x}^T P \tilde{x} + \tilde{h}^T P \tilde{h}
\]

(9)

We can now formulate a quadratic programming version of the problem (7) where the scenario information is contained by the covariance matrix, \( Q \in \mathbb{R}^{2N \times 2N} \), and the vector \( q \in \mathbb{R}^{2N \times 1} \), contains expected values for each tradable instrument. By appending the transaction cost vectors into \( c = [c_p, c_h] \), we formulate the stochastic programming problem (SP) as

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} x^T Q x + q^T x + b + \alpha c^T x \\
\text{subject to} & \quad x \geq 0.
\end{align*}
\]

(10)

We note that (10) is a convex model since \( Q \) is a positive (semi) definite matrix. Hence, any local optimum found will also be a global optimum. Solving (7) with a stochastic programming interior point solver such as the one developed by Blomvall and Lindberg (2002) that require \( n_k \) iterations would have a computational complexity \( O(\eta_i SN^2 + N^3) \) and through the reformulation (10) it is instead \( O(SN^2 + n_i N^3) \). With \( S \gg N \), precalculation of scenario effects leads to significantly reduced computational times.

3.3. Estimating the perceived transaction costs

The details of the transaction costs used in the optimization problem have been left out until now. As stated before, the bid-ask spreads must be crossed when trading with another counterparty. However, neither transaction costs nor other values are immediately exchanged when entering a vanilla swap contract. The cost of entering such a contract with a fixed interest rate is given by the NPV, which depends on the theoretical pricing method. An exact interest rate interpolation method replicates all the contracts exactly, implying zero cost for all instruments used to construct the term structure. Such methods are sensitive to noise in quoted prices and may, therefore, imply unrealistic forward rates. The term structures by Blomvall (2017) are more realistic, which should translate into better estimates of the NPV. Since the method uses a trade-off between price errors and roughness, the costs implied by NPV will be non-zero. To emphasize that these costs are dependent on the term structure measurement method, these are denoted perceived transaction costs. Figure 2 shows that these perceived costs (\( \Delta \tilde{f}_t \text{tantime} \) and \( \Delta \tilde{f}_t \text{theta} \)) accurately explain what drives the growth in portfolio value and can therefore be viewed as a fair proxy for the unobservable costs.

The NPV of the swaps is calculated using both the quoted bid and ask yields for the fixed leg. By using the pricing function, \( P_t \), for instrument \( i \), at time \( t \), with term structure \( f_t \) and quoted yields, \( y_{bid} \) and \( y_{ask} \), the perceived transaction costs for the paying and receiving positions are estimated as

\[
c_{i,t,p} = \max \{ 0, -P_t(f_t, y_{ask}) \}
\]

(11)

\[
c_{i,t,r} = \max \{ 0, -P_t(f_t, y_{bid}) \}.
\]

Since the Blomvall (2017) model allows instrument yields to deviate from the mid yields, we will get different transaction costs for entering paying and receiving positions. The yield deviations will usually be inside the bid-ask spread, but sometimes, contracts are priced outside this spread. For such negative perceived costs, the stochastic programming model (10) would accept an increase in variance. The max-operator ensures that the objective to hedge risk remains intact. Note that a negative cost does not imply an arbitrage opportunity as entering such a contract increases the theoretical portfolio value, but may lead to increased losses in some scenarios. Traditional term structure interpolation methods commonly create term structures which are not plausible. Rather than trusting these completely, it is more likely that the prices are affected by noise (indicative prices, asynchronous prices, tick-size, fat-finger errors, etc.) and that this should be considered in the optimization model. As the performance attribution shows, this price difference explains the rapid increase in value, confirming that the modeling is appropriate.\(^6\)

4. Scenario generation

By computing daily innovations of forward rates, \( \Delta f_t = f_t - f_{t-1} \), we use PCA to decompose the covariance matrix of daily innovations into \( C = E \Lambda E^T \). The risk factors loadings are given by the eigenvector matrix, \( E \). The diagonal matrix, \( \Lambda \), consisting of the corresponding eigenvalues also constitutes the covariance matrix for the orthogonal risk factors. By using a subset consisting of the \( k \) risk factors with highest variance (eigenvalues), we can approximate the covariance matrix, \( C \), as \( \hat{C} = E \Lambda_{k} E^T \). As shown by

---

\(^6\) Note that negative costs successfully explain the growth in portfolio value, even though the SF-model cannot see these gains.
Hagenbjörk & Blomvall (2019), six risk factors are sufficient to explain almost all in-sample forward rate variance across all maturities. The principal components for the risk factors can be computed as \( \xi = \mathbf{E}_k ^T \Delta \mathbf{f} \). These risk factors will be uncorrelated in sample, having a diagonal covariance matrix, \( \mathbf{A}_k \). By drawing random numbers from the multivariate normal distribution \( \hat{\xi}_t \sim N(0, \mathbf{A}_k) \), we can easily obtain forward rate scenarios by computing the changes in the term structure as

\[
\Delta f_t = \mathbf{E}_k \hat{\xi}_t, \quad \hat{\xi}_t \sim N(0, \mathbf{A}_k).
\]

Instead of using the constant volatility from the in-sample period, the empirical variance of each risk factor is used as all models are recalibrated at the beginning of each year using historical data from the most recent six years. The scenarios are generated by antithetic sampling.

Generating thousands of scenarios using only six risk factors for the 2N investment decisions introduces linear dependencies that would be exploited by the SP model. PCA does not explain all variance in the covariance matrix. The missing unexplained variance, \( \nu_{\text{unexp.}} = \text{diag}(\mathbf{C}) - \text{diag}(\mathbf{E}_k ^T \mathbf{A}_k \mathbf{E}_k) \), is added as normally distributed noise \( \Psi \sim N(0, \text{diag}(\nu_{\text{unexp.}})) \). This amount of added noise proves to be insufficient to maintain long-term portfolio stability for the more risk-averse SP models with low values of \( \alpha \). We, therefore, add additional noise with a constant mean across the entire maturity spectrum according to \( \Psi \sim N(0, \text{diag}(\nu_{\text{unexp.}} + \nu)) \). We find that a level of noise that corresponds to a spot rate volatility of \( \nu = 0.01 \) provides the lowest portfolio variance over the in-sample period. In comparison, the average spot rate volatility for the in-sample period is 0.13%. Similarly to the risk factor variance, we also scale the amount of noise by the ratio between the current term structure variance and the variance observed during the in-sample period.

5. Simulation implementation

We use a custom-built extension to the C++ framework QuantLib (Ametrano & Ballabio, 2000–2019) to handle the backtest. The trading process starts with an empty portfolio, and customer trades and hedging are carried out for each day of the 4166 out-of-sample days. We separate the customer part of the portfolio, which is shared among all models. At the end of the simulations, this portfolio contains 10,734 instruments and has handled 1,468,814 transactions in the form of trades, matured instruments, cash flow separations, and subsequent cash flow payments. Each day, all portfolios must be valued under the term structure scenarios. In order to carry out all valuations in a reasonable time, the cash flows of all swaps in a portfolio are aggregated into one composite instrument. Multiple hedging methods can be simulated simultaneously, which will make the simulation faster per model since a lot of the tasks for the customer portfolio only need to be performed once.

The SP models are solved using 5000 scenarios which can be shared among all models that use the same simulation method. Sharing scenarios leads to portfolio valuations under the scenarios consuming most of the time, but this task could theoretically be parallelized. Simulating 20 models over a single date takes around one minute when the portfolio has reached its full size.

6. Data

The data used for estimating term structures in this study has been retrieved from Thomson Reuters EIKON. We focus on the US market due to its high liquidity and the availability of data. The data consist of bid-ask yields for the US 3-month LIBOR, all available forward rate agreements, and the 1–10-year interest rate swaps. All these instruments are used as input to the Blomvall (2017) method for measuring term structures, while the bootstrapped curves only contain the 3-month LIBOR, the 3 × 6 and 6 × 9 months forward rate agreements, and the 1–10 year swaps. Since the yields acquired from EIKON contain some incorrect values, we perform some basic cleaning of the data. This process is described by Blomvall & Hagenbjörk (2019), where the same data set is used.

7. Empirical results

The hedging backtest is carried out using the simple normally distributed scenario-generation methods (12). To span a frontier of possible trade-offs between volatility and daily earnings, we use various hedging parameters, \( \alpha \), spread out evenly in logarithmic space. We present the results in Fig. 4, where the traditional models and the naïve back-to-back hedge are included for reference. All the four reference models are below the frontier spanned by the stochastic programming model. This implies that we can either obtain a lower risk with the same average daily return or a higher average daily return to the same risk. The figure contains 95% confidence intervals for the point estimates of the daily earnings. As the methods can be compared pairwise, these confidence intervals should not be used to compare specific pairs of models. As shown in the upper right triangular part of Table 3, pairwise t-tests show that all SP models have resulting daily earnings that are significantly higher than Bumping. When testing for differences in volatilities between models in the lower-left triangular part of Table 3, the results in Fig. 4 are confirmed except for the difference in volatility between \( \alpha = 10^{-4} \) and Bumping, which is insignificant.

In Table 2, we present a detailed performance attribution of the results of some of the SP models. We see that SP with \( \alpha = 10^{-4} \) produces lower volatility than Bumping with a slightly lower transaction cost ($3679 compared to $3683), which results in a higher daily earning that is statistically significant. According to the results for \( \alpha = 10 \) in Table 2, we find that it would be possible to reduce the portfolio volatility by an order of magnitude (88.0 to 6.9) while gaining value from the estimated transaction costs of the hedge positions. If we study the performance attribution of this model more closely in Fig. 5, we can confirm that value is gained from hedging costs, especially during the turbulent times of the financial crisis. We see that the magnitude of the risk factors is larger now when the portfolio is not always fully hedged. However, no risk factor dominates the risk and the risk that cannot be explained by our six risk factors, \( e_t^j \), also stays moderate. To avoid the suspicion that our theoretical pricing function fictitiously creates this excess portfolio value through the means of the term structure measurement method, we have included the portfolio value measured with cubic spline term structures (red line in Fig. 5). As can be seen, we do not systematically overvalue or undervalue the portfolio in comparison to the more volatile cubic spline valuation.

The customer trades are exogenous and the out-of-sample time period 2002–2018 contains periods with varying volatility, which implies that the characteristics of the portfolio change exogenously over time. By weighting cost against risk, the SP-model can dynamically adapt to the everchanging conditions. This would not be the case if risk or transaction costs were included as constraints that are independent of the exogenous variables.

All traditional hedging methods can eliminate almost all the risk exposure (Table 1). The volatility for Bumping is 0.02 and the average daily return is 0.2189 (Table 1), and for the SP-model with \( \alpha = 10 \) it has increased to 1.273 (Table 2), an increase of 480%.
The t-statistic that Bumping has higher average return than the SP-model with \( \alpha = 10 \) is \(-9.62 \) (Table 3). As shown in Fig. 5, the cubic spline valuation has high negative autocorrelation \((-0.46)\). Even though our valuation is much more stable, it still gives a large negative autocorrelation \((-0.32)\). The negative autocorrelation is likely to be an artifact of market price noise, rather than an indication that simple technical analysis could be useful. With access to even better term structure measurement methods, the volatility in gains (6.9) would be much smaller, and the t-statistic would be an even larger negative value. The volatility computed from the scenarios varies due to the exogenous variables. The volatility always succeeds 23, with an average volatility of 8.9, which is close to the realized volatility (6.9). The SP-model can give even smaller risks compared to traditional methods, but also find cost efficient hedges (Fig. 4), where the optimal decision depends on the risk aversion captured by \( \alpha \).
7.1. Stochastic dominance

Stochastic dominance can be utilized when determining a preferable distribution. We consider the outcome of the daily earning as a stochastic random variable. The simplest and strongest case of stochastic dominance is statewise stochastic dominance. Method A is statewise dominant over method B if A gives at least as good a result in every possible state and a strictly better result in at least one state. Since we do not have access to the outcome in all possible states, we turn to weaker formulations. First-order dominance implies that the distribution functions are not allowed to cross each other. Let $F_A$ and $F_B$ be distribution functions for methods A and B. Method A dominating B in first-order, denoted by $F_A \succ F_B$, is given by the following definition.

**Definition 1.** Method A has first-order stochastic dominance over method B if and only if $F_A(x) \leq F_B(x) \forall x$, with strict inequality for some $x$.

Second-order stochastic dominance is weaker than first-order dominance since it requires the investor to be risk-averse. For the case of hedging, this is acceptable since hedging only would be of interest to a risk-averse investor in the first place.

**Definition 2.** A dominates B in the second-order, denoted by $F_A \succ_2 F_B$ if and only if

$$\int_{-\infty}^{x} F_A(t) \, dt \leq \int_{-\infty}^{x} F_B(t) \, dt$$

$\forall x$ with strict inequality at some $x$.

Davidson & Duclos (2000) define a stochastic dominance of any order $s$. Let $F^x_A(x) = F(x)$ and

$$F^x_A(x) = \int_0^x F_A^{(s-1)}(y) \, dy,$$  \hspace{1cm} (14)$$

for any integer $s \geq 2$, and define $F^s_B(x)$ accordingly. Method A is said to dominate method B at order $s$ if $F^x_A(x) \leq F^x_B(x)$ for all $x \in \mathbb{R}$. Davidson & Duclos (2000) use this definition to formulate a statistical test of stochastic dominance of order $s$, henceforth called DD. Let $X$ and $Y$ be two random variables with distribution functions $F_A$ and $F_B$, which in this application correspond to the daily earnings for two hedging methods from the $N=4166$ observations, $x_i$ and $y_i$. To test for stochastic dominance, the distributions should be examined for all $x$ in the support of $F_A$ and $F_B$. However, this is empirically impossible, and thus we test the hypothesis on a finite number of points $z_i$. We test the null hypothesis that $F_A$ and $F_B$ are equal,

$$H_0 : F_A = F_B,$$

against the following three hypotheses

$$H_{IA} : F_A \succ_1 F_B,$$

$$H_{IB} : F_A \succ_2 F_B,$$

$$H_1 : F_A \not\equiv F_B.$$

The null hypothesis is only rejected if the largest test statistic $\max_i T_i(z_i)$ is significant. The test’s significance level can be determined using the studentized maximum modulus (SMM) distribution, which provides a conservative statistic when the test statistics $T_i(z_i)$ are independent.
To obtain the values $z_i$ used for the test statistic and computation of the test statistics, we use the technique described by Fang, Wong, & Lean (2005) and Lean, Wong, & Zhang (2008). The sample points $x_i, y_i, i = 1, \ldots, N$ are sorted and divided into $M$ major intervals, which are further uniformly divided into $m$ minor intervals. This ensures non-omission of important information between the major grid points. Using the critical value for the SMM distribution for $M$ points instead of $Mm$ helps to satisfy the independence assumption to achieve a conservative test. We use $M = m = 10$ as suggested by Fang et al. (2005) and Lean et al. (2008). The critical values are tabulated in Stoline & Ury (1979).

Lean et al. (2008) compare the performance of three commonly used statistical tests for Stochastic Dominance. The comparison is carried out using Monte Carlo simulations to examine statistical power and size, i.e., the probability of making a Type I error by falsely rejecting the null hypothesis. The tests are compared for correlated and heteroskedastic data, which are properties inherent in the time series of daily earnings. Since the daily differences in each model result from the same movements in risk factors for each day, we will have some correlation even though the exposure of different hedges will differ. Lean et al. (2008) find all studied methods to be conservative, but the DD to perform best in terms of both size and power for correlated distributions. For heteroskedastic data, they experience a significant reduction in power. However, the power of the DD test remains reasonably good for large sample sizes. They also conclude that the DD test is conservative even for correlated distributions and for second- and third-order stochastic dominance under a heteroskedastic process.

We start by comparing SP $\alpha = 10^{-4}$ to Bumping, illustrated in the left panel of Fig. 6. The two overlapping lines $\hat{F}_A(z)$ and $\hat{F}_B(z)$ represent the empirical cumulative distribution functions for the two methods and the blue line, $\hat{F}_B^2(z) - \hat{F}_A^2(z)$, approximates the integral of their difference. This line, which turns red when dropping below zero, is plotted on a different scale. We see that the blue line does decline at several places, indicating that the empirical distribution functions cross each other multiple times. Since we have a limited number of observations, $X$ and $Y$, we cannot draw any conclusions from this alone since the DD test allows for $\hat{F}_B^2(z) - \hat{F}_A^2(z)$ to be negative at some points as long as the variance in the denominator is large enough. After performing the DD test, we accept the null hypothesis $H_0$ that we cannot distinguish between the two methods in the first order ($s = 1$). The blue line staying above zero over the entire empirical distribution indicates the potential for but does not guarantee second-order stochastic dominance. To accept the hypothesis $H_{1\alpha}$ that SP $\alpha = 10^{-4}$ dominates Bumping in the second-order, $\hat{F}_B^2(z) - \hat{F}_A^2(z)$, must be large enough in relation to the variance for every one of the investigated points. Performing the DD test for $s = 2$ confirms that SP $\alpha = 10^{-4}$ indeed dominates Bumping at significance level 99%. This result means that any risk-averse investor would prefer to use SP $\alpha = 10^{-4}$ over Bumping when setting up a hedge for a portfolio over the time frame of one single day. This is a very strong result.

When comparing SP $\alpha = 10^{-2}$ to Bumping, illustrated in the middle panel of Fig. 6, we see that the empirical distribution for SP $\alpha = 10^{-2}$ is above that of Bumping for the lowest outcomes of day-to-day changes in portfolio value. This is natural since it accepts a higher risk and thereby will experience higher losses over some days. Even though it makes up for the losses on average, it is clear that this strategy will not dominate Bumping since $\hat{F}_B^2(z) - \hat{F}_A^2(z)$ declines for $z_i$'s in the left tail. The test confirms this, and hence, SP $\alpha = 10^{-2}$ will not be preferred by all risk-averse investors for hedging over a one-day time horizon. However, hedging a swap book is not a task just carried out over a single day. Instead, it is a process that is carried out repeatedly over long periods. Therefore, we also compare these methods for a weekly time frame using non-overlapping 5-business-day periods, displayed in the right panel of Fig. 6. The DD-test confirms that the SP $\alpha = 10^{-2}$ method is preferred by all risk-averse investors, at significance level 99%, when evaluating on a weekly basis. Given the increasing portfolio values in the frontier in Fig. 4, and by extending the argument of increased time frames, it is possible to statistically show that a method with higher hedging parameter $\alpha$ dominates all methods with lower $\alpha$, by choosing the right time frame. For $\alpha = 1$ to dominate Bumping in the second-order, a bi-weekly time frame (10 business days) is needed, and for $\alpha = 10$ to dominate Bump-
ing at the second-order, a monthly time frame (20 business days) is needed.

8. Discussion

In practice, it is impossible to find optimal decisions for most real-life problems. This property enables us to determine improvements through systematic application of operational research. The improvement of delta-hedging a bank’s interest rate swap book started with (i) improving the optimization model (inverse problem) of measuring the term structure of interest rates (Blomvall, 2017). This allowed for (ii) refined systematic risk factors to be identified (Blomvall & Ndengo, 2013), which was used to improve both (iii) risk measurements (Hagenbjörk & Blomvall, 2019) and (iv) performance attribution (Blomvall & Hagenbjörk, 2019).

An optimal hedge balances the risk against the associated costs. The improved measurement of the ex-ante risk (iii) and the expected realized costs and losses due to risk factors (iv) have been fundamental to (v) formulate the improved stochastic programming model (7). The tools (i-iv) make it possible to determine why a hedging decision cannot be optimal. They also make it possible to identify properties with limited impact. The formulation and implementation of the stochastic programming model (7) is the result of manual meta-optimization, where misconceptions and significant bugs have been eliminated iteratively using (i-iv).

The impossibility of finding optimal decisions derives from the complexity of the real world and the infinite number of possible mathematical approximations to describe it, where only a minuscule share leads to improved decisions. The implementation process has provided us with a glimpse of the vast number of possible formulations that do not improve decisions. It is reassuring that the final formulation leads to such persistent improvements as the underwater plot relative to delta hedging shows in Fig. 3, which results in high significance in the statistical tests in Table 3 and second-order stochastic dominance in Section 7.1. These results increase the likelihood that an improved formulation has been found and that the associated costs with delta hedging seen in Fig. 2 are much larger than necessary (compare to Fig. 5). Since delta hedging neither considers the significance of the risk factors nor the associated costs to reduce risk, this inevitably leads to unnecessarily high costs.

This study is limited to a single-curve setting as a large data set is necessary to validate the model. The market has progressed toward multi-curve sets with OIS discounting, and more recently to instruments based on interest rates such as ESTR but these have relatively short data sets. In principle, the delta-hedging decisions for these cases can be improved using the same operational research methods (i-v).

9. Conclusions

In this paper, we propose a stochastic programming model to hedge interest rate risk while accounting for transaction costs. We formulate an optimization problem that weighs reduced variance against perceived transaction costs through a hedging parameter, \( \alpha \). The Blomvall (2017) framework is used for measuring the term structure of forward rates. These high-quality term structures enable us to obtain accurate measurements of the systematic risk factors, as well as the perceived transaction costs. The hedging is carried out for different values of \( \alpha \) using a realistic simulation setting between 2002 and 2018. Our results are compared to a naïve method and the traditional hedging methods by Hagan & West (2006). Utilizing the residual-free performance attribution framework presented by Blomvall & Hagenbjörk (2019), we are able to study the results of each model in detail. Statistical tests confirm that our SP model outperforms the traditional methods in terms of lower volatility and higher portfolio values. Using statistical tests for second-order stochastic dominance, we show that the proposed SP model is preferred by all risk-averse investors when hedging a portfolio over a single day. We find that a simple scenario-generation method with normally distributed risk factor innovations is sufficient to produce good results in this hedging setting. However, it is crucial to provide uncorrelated noise to prevent the optimization model from taking advantage of the reduced dimensionality in the scenario set. The main contribution of the paper is a flexible hedging model that is preferred over the traditional methods by all risk-averse investors. The proposed model is thus a step towards an optimal hedging policy.

Funding

This research was funded in part by Sweden’s Innovation Agency (Grant no. 2019–04798).

References


