7-2021

A unified market model for swaptions and constant maturity swaps

Chyng Wen TEE
Singapore Management University, cwtee@smu.edu.sg

Jeroen KERKHOF

Citation

TEE, Chyng Wen and KERKHOF, Jeroen. A unified market model for swaptions and constant maturity swaps. (2021). *International Journal of Theoretical and Applied Finance*. 24, (4), 1-31.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/6868

A Unified Market Model for Swaptions and Constant Maturity Swaps

Chyng Wen Tee*† Jeroen Kerkhof‡§

Abstract

Internal-rate-of-return (IRR) settled swaptions are the main interest rate volatility instruments in the European interest rate markets. Industry practice is to use an approximation formula to price IRR swaptions based on Black model, which is not arbitrage-free. We formulate a unified market model to incorporate both swaptions and constant maturity swaps (CMS) pricing under a single, self-consistent framework. We demonstrate that the model is able to calibrate to market quotes well, and is also able to efficiently price both IRR-settled and swap-settled swaptions, along with CMS products. We use the model to illustrate the difference in implied volatilities for IRR-settled payer and receiver swaptions, the pricing of zero-wide collars and in-the-money (ITM) swaptions, the implication on put-call parity, and the issue of negative vega. These findings offer important insights to the ongoing reform in the European swaption market.

Keywords: *interest rate market, swaptions, constant maturity swaps, derivative valuation, stochastic volatility models, fixed income market, interest rate models.*

*Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road #05-01, Singapore 178899.
†Corresponding author: Tel: +65 6828 0819. Email: cwtee@smu.edu.sg.
‡Vrije Universiteit Brussel, Pleinlaan 5, 1050, Elsene. VAR Strategies BVBA, Belgium.

§The authors would like to acknowledge the comments and suggestions from an anonymous referee, which lead to much improvement in the flow and clarity of this paper.

1 Introduction

European interest rate markets trade internal-rate-of-return (IRR) settled swaptions as the main interest rate volatility instrument. In the market only at-the-money (ATM) straddles and out-of-the-money (OTM) payers and receivers are liquidly quoted to be usable as calibration instruments. Unlike the dollar market in the USA which trades swap-settled swaptions, and can be readily priced by a closed-form formula via a Black model (Black 1976 and Jamshidian 1997), simplifying assumptions are required for the euro, sterling, and Swiss franc markets in Europe, which are not arbitrage-free. Nevertheless, the difference between IRR-settled and swap-settled swaptions is often incorrectly considered to be minor, hence an approximation formula based on Black model is generally used to price IRR-settled swaptions across the industry. This pricing framework is not self-consistent, and Mercurio (2008) and Henrard (2010) have established mathematically that the standard approximation formula is not arbitrage-free when applied to IRR-settled swaptions.

IRR-settled swaptions are settled in cash, based on the value of the payoff observed on the maturity date. For this reason IRR-settled swaptions in the European markets are also commonly referred to as "cash-settled swaptions". However, we point out that swap-settled swaptions can also be settled either via a physical swap or by cash. In fact, it is common for swap-settled swaptions in the USD market to be cash-settled. Hence, in this paper, we adopt the naming convention of "IRR-settled" and "swap-settled" for greater clarity.

After the European swaption market reform in 2018[1], market makers have been quoting both swap-settled and IRR-settled swaptions, alongside zero-wide collars and constant maturity swaps (CMS) products, without a unified model to price all of these instruments within a single framework. Over the last few years, the volume of traded swap-settled swaptions in the euro market has steadily increased. Today, approximately 80% of newly traded swaptions are already of the swap-settled type. Going forward, the volume of IRR-settled swaptions will likely continue to decline. Nevertheless, there is a large volume of existing IRR-settled swaptions which will still need to be valued and risk-managed until their expiry.

The standard approach to swaption and CMS pricing is to first formulate a market model for swap-settled swaptions before extending it to handle IRR-settled swaptions by means of an approximation formula. This extension involves arbitrage-prone assumptions. CMS products are then priced by performing convexity correction or static replication.

Two major simplifying assumptions are required to arrive at the approximation formula used widely in the industry to price IRR-settled swaptions. First, the concave and convex nature of the IRR-settled payer and receiver swaption payoffs, respectively, are ignored when evaluating the expectation of the swaption payoffs. Second, the expectations are evaluation under a risk-neutral measure that is not associated to the chosen numeraire (Section 2 for further discussions). Making these two simplification allows one to recover the standard Black formula for IRR-settled swaption pricing. There is a strong preference among practitioners to use a market model based on Black formula to price swaptions in a quick and efficient manner, since swaptions serve as the basis to value exotic interest rate volatility products. The standard market model used in the swaption market is the stochastic alpha-beta-rho (SABR) model (Hagan et al. 2002), which provides an analytical expression for the implied Black volatility. This can then be readily substituted to the Black formula to obtain the swaption price. Although this approach is exact and accurate for swap-settled swaptions, it is only an approximation when applied to IRR-settled swaptions due to the simplification involved.

The approximation formula based on Black model used in the industry leads us to a number of incorrect conclusions in terms of IRR-settled swaption pricing (Tee & Kerkhof 2014). First, according to the approximation formula, the IRR-settled ATM payers and ATM receivers will have identical price, which leads to the zero-wide collar pricing problem. Second, the approximation formula cannot distinguish the difference between the implied volatilities for IRR-settled payers and receivers, which in turn leads to the in-the-money (ITM) swaptions pricing problem. Third, based on the approximation formula, an increase in volatilities always lead to an increase in swaption prices. In reality, this is not always the case, which lead to the negative vega problem. We will demonstrate in this paper that a unified market model can be formulated to address and resolve all three issues listed here in a consistent manner.

Moreover, market has been quoting zero-wide collars struck at the forward swap rates at a non-zero premiums since 2015, which is incompatible with the approximation formula. An important work Cedervall & Piterbarg (2012) triggered the transformation of how industry prices these products. Both

Lutz (2015) and Feldman & Portheault (2017) investigate the pricing problem of zero-wide collars and realize that for IRR-settled swaptions, zero-wide collars worth should not be worth zero in general. As Pietersz & Sengers (2017) point out, apart from full-fledged term-structure models, a simple arbitrage-free model to consistently value IRR-settled swaptions has been lacking so far. This lack of an efficient market model to price IRR-settled swaptions consistently has impacted market liquidity. For instance, Lutz (2015) points out that liquidity for euro ITM IRR-settled swaptions virtually disappeared due to uncertainty about the proper pricing of those products. He also examines an alternative pricing approach in the form of a terminal swap rate model. On the other hand, Pietersz & Sengers (2017) try to solve this problem by postulating that the swap-annuity numeraire can be modeled as the cash-annuity formula evaluated at a newly introduced discount forward swap rate—this additional degree-of-freedom introduced is used to calibrate to market prices. Another alternative approach explored by Bermin & Williams (2017) is to formulate an arbitrage-free pricing model within the Markov functional modeling framework to tackle the pricing problem of IRR-settled swaptions and CMS. They also apply their model to investigate the extent of convexity adjustment and forward sensitivity. The importance of change of measure as the main component for consistent market price formation has been further studied in Feldman (2020).

In this paper, we formulate a unified pricing framework in the form of a displaced-diffusion stochastic volatility model to handle swap-settled and IRR-settled swaptions along with CMS products in a consistent and arbitrage-free manner. We demonstrate how it can be calibrated to observable market swaptions and CMS product quotes. Our framework is able to handle the whole range of swap and CMS markets derivatives without any arbitrary adjustment or simplification. To ensure efficiency, we use an expansion method to obtain semi-analytical pricing formulae.

Using our unified market model, we demonstrate that IRR-settled ATM payers and receivers are not worth the same, and zero-wide collar can be efficiently priced. We prove that payer and receiver implied volatility curves ought to be separated—in other words, one cannot use the OTM implied volatilities quoted in the market to value ITM IRR-settled swaptions via the approximation formula, thereby resolving the ITM swaption pricing problem. We also use our unified framework to elaborate how the two simplifying assumptions made in deriving the approximation formula used in the market lead to offsetting errors, resulting in a smaller overall error. This explains why the market has been able to use the approximation formula without incurring large discrepancies. Finally, we demonstrate the existence of negative vega in IRR-settled swaptions.

This paper is organized as follows: Section 2 presents an overview of the swap market model, the approximation and simplifying assumptions required to handle IRR-settled swaptions, and how convexity correction for CMS can be calculated by a replication approach. In Section 3, we formulate a unified market model for both swaptions and CMS markets. We also derive semi-analytical pricing formulae for our model based on an expansion method. We present a number of key analyses and results in Section 4, where numerous important insights including the pricing of ITM swaptions and zero-wide collars, the implication on put-call parity, and negative volatility sensitivities are expounded. Finally, conclusions are drawn in Section 5.

2 Swap Market Model

Market models can be defined as the family of models that takes observable market interest rates, for instance the swap rate or the LIBOR rate, as the basis for modeling. The main advantage of using this approach is that standard pricing formulae based on Black model can be used to price liquid vanilla instruments. Market models generally postulate a geometric Brownian motion for the market rates under consideration. This approach is formalized after the LIBOR Market Model (LMM) (Miltersen, Sandmann & Sondermann 1997 and Brace, Gatarek & Musiela 1997) and the Swap Market Model (SMM) (Jamshidian 1997) were introduced.

Although it is possible to price swap- and IRR-settled swaptions, along with CMS products, in a consistent and arbitrage-free framework using a short rate model like the Hull-White model (Hull & White 1990) or a term structure model like the Heath-Jarrow-Morton model (Heath, Jarrow & Morton 1992), practitioners prefer to use market models for liquid volatility instruments, given that market models are generally more straightforward to calibrate and price, and can fit the volatility smile profiles better. Moreover, market model avoid having to work on unobservable instantaneous spot or forward interest rate processes, and is therefore more intuitive.

2.1 Market Quoting Convention

The swaption market convention is to quote at-the-money (ATM) straddles along with out-of-the-money (OTM) payers and receivers across a range of strikes — typically ATM ± 300 bps (basis points), with 50 bps spacing between adjacent strikes. The ATM strike is defined as the value of the forward swap rate evaluated under the swap annuity risk-neutral measure. As Jamshidian (1997) points out, the natural martingale measure of the swap rate is the risk-neutral measure associated with the swap annuity numeraire. Consequently, the ATM point is by definition always the expected forward swap rate evaluated under the swap annuity measure. The quotes provided by swaption brokers can be either premium or implied volatility, calculated based on a pre-determined discounting convention (frequently overnight index swap (OIS)). After the 2007-2009 Global Financial Crisis, market has switched to OIS discounting for collateralized trades (Bianchetti & Carlicchi 2013). In-the-money (ITM) quotes—low strike payers and high strike receivers—are not readily available, and need to be priced independently by the trading desks. For swap-settled swaptions, the ITM payers and receivers can be priced using the same implied volatility as the OTM swaption of the same strike and maturity. However, this is not the case for IRR-settled ITM swaptions. We will explain in subsequent sections that IRR-settled ITM swaptions need to be determined using a pricing model. Given that IRR-settled swaptions are the main interest rate volatility instruments in the European markets, this has important implication for institutional traders holding large swaption portfolio, as their daily interest rate portfolio mark-to-market will include a large number of ITM swaptions, whose prices are model-dependent.

Let K denote the strike of a swaption, and let $V_p(K)$ and $V_r(K)$ denote the price of a receiver and payer swaptions struck at K, respectively. Exhibit I tabulates the liquid market instruments in the EUR swaption market that can be used for model calibration. There are brokers screens for ATM swaptions' implied normal volatilities of both swap- and IRR-settlement types (Panel A). For volatility smile, there are broker quotes for for ATM-200bps to ATM+200bps across different expiry-tenor pairs (Panel B). In addition to these, there are also forward premium quotes for zero-width collars and straddles for IRR-settled swaptions (Panel C). Overall there is a rich set of liquid swaption market data available for model calibration.

Exhibit I: Quoting Convention in the EUR Swaption Market

Panel A: ATM Swaptions (Normal Volatilities)

	Tenors										
Expiries	$1Y$	$2Y$	$3Y$	$4Y$	$5Y$	$\cdots$	$10Y$	$15Y$	$20Y$	$25Y$	$30Y$
$1M$	σ_{1m1y}	σ_{1m2y}	σ_{1m3y}	σ_{1m4y}	σ_{1m5y}	$\cdots$	σ_{1m10y}	σ_{1m15y}	σ_{1m20y}	σ_{1m25y}	σ_{1m30y}
$2M$	σ_{2m1y}	σ_{2m2y}	σ_{2m3y}	σ_{2m4y}	σ_{2m5y}	$\cdots$	σ_{2m10y}	σ_{2m15y}	σ_{2m20y}	σ_{2m25y}	σ_{2m30y}
$3M$	σ_{3m1y}	σ_{3m2y}	σ_{3m3y}	σ_{3m4y}	σ_{3m5y}	$\cdots$	σ_{3m10y}	σ_{3m15y}	σ_{3m20y}	σ_{3m25y}	σ_{3m30y}
$6M$	σ_{6m1y}	σ_{6m2y}	σ_{6m3y}	σ_{6m4y}	σ_{6m5y}	$\cdots$	σ_{6m10y}	σ_{6m15y}	σ_{6m20y}	σ_{6m25y}	σ_{6m30y}
$9M$	σ_{9m1y}	σ_{9m2y}	σ_{9m3y}	σ_{9m4y}	σ_{9m5y}	$\cdots$	σ_{9m10y}	σ_{9m15y}	σ_{9m20y}	σ_{9m25y}	σ_{9m30y}
$1Y$	σ_{1y1y}	σ_{1y2y}	σ_{1y3y}	σ_{1y4y}	σ_{1y5y}	$\cdots$	σ_{1y10y}	σ_{1y15y}	σ_{1y20y}	σ_{1y25y}	σ_{1y30y}
$18M$	σ_{18m1y}	σ_{18m2y}	σ_{18m3y}	σ_{18m4y}	σ_{18m5y}	$\cdots$	σ_{18m10y}	σ_{18m15y}	σ_{18m20y}	σ_{18m25y}	σ_{18m30y}
$2Y$	σ_{2y1y}	σ_{2y2y}	σ_{2y3y}	σ_{2y4y}	σ_{2y5y}	$\cdots$	σ_{2y10y}	σ_{2y15y}	σ_{2y20y}	σ_{2y25y}	σ_{2y30y}
$3Y$	σ_{3y1y}	σ_{3y2y}	σ_{3y3y}	σ_{3y4y}	σ_{3y5y}	$\cdots$	σ_{3y10y}	σ_{3y15y}	σ_{3y20y}	σ_{3y25y}	σ_{3y30y}
$4Y$	σ_{4y1y}	σ_{4y2y}	σ_{4y3y}	σ_{4y4y}	σ_{4y5y}	$\cdots$	σ_{4y10y}	σ_{4y15y}	σ_{4y20y}	σ_{4y25y}	σ_{4y30y}
$5Y$	σ_{5y1y}	σ_{5y2y}	σ_{5y3y}	σ_{5y4y}	σ_{5y5y}	$\cdots$	σ_{5y10y}	σ_{5y15y}	σ_{5y20y}	σ_{5y25y}	σ_{5y30y}
$6Y$	σ_{6y1y}	σ_{6y2y}	σ_{6y3y}	σ_{6y4y}	σ_{6y5y}	$\cdots$	σ_{6y10y}	σ_{6y15y}	σ_{6y20y}	σ_{6y25y}	σ_{6y30y}
$7Y$	σ_{7y1y}	σ_{7y2y}	σ_{7y3y}	σ_{7y4y}	σ_{7y5y}	$\cdots$	σ_{7y10y}	σ_{7y15y}	σ_{7y20y}	σ_{7y25y}	σ_{7y30y}
$10Y$	σ_{10y1y}	σ_{10y2y}	σ_{10y3y}	σ_{10y4y}	σ_{10y5y}	$\cdots$	σ_{10y10y}	σ_{10y15y}	σ_{10y20y}	σ_{10y25y}	σ_{10y30y}
$12Y$	σ_{12y1y}	σ_{12y2y}	σ_{12y3y}	σ_{12y4y}	σ_{12y5y}	$\cdots$	σ_{12y10y}	σ_{12y15y}	σ_{12y20y}	σ_{12y25y}	σ_{12y30y}
$15Y$	σ_{15y1y}	σ_{15y2y}	σ_{15y3y}	σ_{15y4y}	σ_{15y5y}	$\cdots$	σ_{15y10y}	σ_{15y15y}	σ_{15y20y}	σ_{15y25y}	σ_{15y30y}
$20Y$	σ_{20y1y}	σ_{20y2y}	σ_{20y3y}	σ_{20y4y}	σ_{20y5y}	$\cdots$	σ_{20y10y}	σ_{20y15y}	σ_{20y20y}	σ_{20y25y}	σ_{20y30y}
$25Y$	σ_{25y1y}	σ_{25y2y}	σ_{25y3y}	σ_{25y4y}	σ_{25y5y}	$\cdots$	σ_{25y10y}	σ_{25y15y}	σ_{25y20y}	σ_{25y25y}	σ_{25y30y}
$30Y$	σ_{30y1y}	σ_{30y2y}	σ_{30y3y}	σ_{30y4y}	σ_{30y5y}	$\cdots$	σ_{30y10y}	σ_{30y15y}	σ_{30y20y}	σ_{30y25y}	σ_{30y30y}

2.2 Pricing Swap-Settled Swaption

In this section, we review the standard market approach of using the SABR model for swap-settled swaptions pricing. We will then point out the discrepancies in applying the same model to price IRR-settled swaptions, and how the approximation formula used widely in the market is derived by making two simplifying assumptions.

A swap-settled payer swaption (denoted here as V_p^{Swp}) can be priced as

$$V_p^{\text{Swp}}(0, K) = \mathbb{E}^* \left[e^{-\int_0^T r_t \, dt} A_p(T)(S(T) - K)^+ \right]. \tag{2.1}$$

Note that receiver swaptions can be priced in the same way. We will use payer swaptions throughout this paper to illustrate the pricing models discussed. Here, r_t is the spot rate at time t, and A_p is the

Panel B: Swaption Volatility Smile (Normal Volatilities)

Tenor-Expiry	-200	-100	-50	-25	ATM	25	50	100	200
$1Y1Y$	$\sigma_{1y1y}^{ATM-200}$	$\sigma_{1y1y}^{ATM-100}$	σ_{1y1y}^{ATM-50}	σ_{1y1y}^{ATM-25}	σ_{1y1y}^{ATM}	σ_{1y1y}^{ATM+25}	σ_{1y1y}^{ATM+50}	$\sigma_{1y1y}^{ATM+100}$	$\sigma_{1y1y}^{ATM+200}$
$3M2Y$	$\sigma_{3m2y}^{ATM-200}$	$\sigma_{3m2y}^{ATM-100}$	σ_{3m2y}^{ATM-50}	σ_{3m2y}^{ATM-25}	σ_{3m2y}^{ATM}	σ_{3m2y}^{ATM+25}	σ_{3m2y}^{ATM+50}	$\sigma_{3m2y}^{ATM+100}$	$\sigma_{3m2y}^{ATM+200}$
$2Y2Y$	$\sigma_{2y2y}^{ATM-200}$	$\sigma_{2y2y}^{ATM-100}$	σ_{2y2y}^{ATM-50}	σ_{2y2y}^{ATM-25}	σ_{2y2y}^{ATM}	σ_{2y2y}^{ATM+25}	σ_{2y2y}^{ATM+50}	$\sigma_{2y2y}^{ATM+100}$	$\sigma_{2y2y}^{ATM+200}$
$1Y5Y$	$\sigma_{1y5y}^{ATM-200}$	$\sigma_{1y5y}^{ATM-100}$	σ_{1y5y}^{ATM-50}	σ_{1y5y}^{ATM-25}	σ_{1y5y}^{ATM}	σ_{1y5y}^{ATM+25}	σ_{1y5y}^{ATM+50}	$\sigma_{1y5y}^{ATM+100}$	$\sigma_{1y5y}^{ATM+200}$
$5Y5Y$	$\sigma_{5y5y}^{ATM-200}$	$\sigma_{5y5y}^{ATM-100}$	σ_{5y5y}^{ATM-50}	σ_{5y5y}^{ATM-25}	σ_{5y5y}^{ATM}	σ_{5y5y}^{ATM+25}	σ_{5y5y}^{ATM+50}	$\sigma_{5y5y}^{ATM+100}$	$\sigma_{5y5y}^{ATM+200}$
$3M10Y$	$\sigma_{3m10y}^{ATM-200}$	$\sigma_{3m10y}^{ATM-100}$	σ_{3m10y}^{ATM-50}	σ_{3m10y}^{ATM-25}	σ_{3m10y}^{ATM}	σ_{3m10y}^{ATM+25}	σ_{3m10y}^{ATM+50}	$\sigma_{3m10y}^{ATM+100}$	$\sigma_{3m10y}^{ATM+200}$
$1Y10Y$	$\sigma_{1y10y}^{ATM-200}$	$\sigma_{1y10y}^{ATM-100}$	σ_{1y10y}^{ATM-50}	σ_{1y10y}^{ATM-25}	σ_{1y10y}^{ATM}	σ_{1y10y}^{ATM+25}	σ_{1y10y}^{ATM+50}	$\sigma_{1y10y}^{ATM+100}$	$\sigma_{1y10y}^{ATM+200}$
$2Y10Y$	$\sigma_{2y10y}^{ATM-200}$	$\sigma_{2y10y}^{ATM-100}$	σ_{2y10y}^{ATM-50}	σ_{2y10y}^{ATM-25}	σ_{2y10y}^{ATM}	σ_{2y10y}^{ATM+25}	σ_{2y10y}^{ATM+50}	$\sigma_{2y10y}^{ATM+100}$	$\sigma_{2y10y}^{ATM+200}$
$5Y10Y$	$\sigma_{5y10y}^{ATM-200}$	$\sigma_{5y10y}^{ATM-100}$	σ_{5y10y}^{ATM-50}	σ_{5y10y}^{ATM-25}	σ_{5y10y}^{ATM}	σ_{5y10y}^{ATM+25}	σ_{5y10y}^{ATM+50}	$\sigma_{5y10y}^{ATM+100}$	$\sigma_{5y10y}^{ATM+200}$
$10Y10Y$	$\sigma_{10y10y}^{ATM-200}$	$\sigma_{10y10y}^{ATM-100}$	σ_{10y10y}^{ATM-50}	σ_{10y10y}^{ATM-25}	σ_{10y10y}^{ATM}	σ_{10y10y}^{ATM+25}	σ_{10y10y}^{ATM+50}	$\sigma_{10y10y}^{ATM+100}$	$\sigma_{10y10y}^{ATM+200}$
$15Y15Y$	$\sigma_{15y15y}^{ATM-200}$	$\sigma_{15y15y}^{ATM-100}$	σ_{15y15y}^{ATM-50}	σ_{15y15y}^{ATM-25}	σ_{15y15y}^{ATM}	σ_{15y15y}^{ATM+25}	σ_{15y15y}^{ATM+50}	$\sigma_{15y15y}^{ATM+100}$	$\sigma_{15y15y}^{ATM+200}$
$10Y20Y$	$\sigma_{10y20y}^{ATM-200}$	$\sigma_{10y20y}^{ATM-100}$	σ_{10y20y}^{ATM-50}	σ_{10y20y}^{ATM-25}	σ_{10y20y}^{ATM}	σ_{10y20y}^{ATM+25}	σ_{10y20y}^{ATM+50}	$\sigma_{10y20y}^{ATM+100}$	$\sigma_{10y20y}^{ATM+200}$
$5Y30Y$	$\sigma_{5y30y}^{ATM-200}$	$\sigma_{5y30y}^{ATM-100}$	σ_{5y30y}^{ATM-50}	σ_{5y30y}^{ATM-25}	σ_{5y30y}^{ATM}	σ_{5y30y}^{ATM+25}	σ_{5y30y}^{ATM+50}	$\sigma_{5y30y}^{ATM+100}$	$\sigma_{5y30y}^{ATM+200}$

Panel C: Zero-Width Collars & Straddles (Premium)

	Forward Premium (cash settlement)									
	Zero-Width Collar					ATM Straddle				
	$2y$	$5y$	$10y$	$20y$	$30y$	$2y$	$5y$	$10y$	$20y$	$30y$
$1y$	V_{1y2y}^{Col}	V_{1y5y}^{Col}	V_{1y10y}^{Col}	V_{1y20y}^{Col}	V_{1y30y}^{Col}	V_{1y2y}^{Col}	V_{1y5y}^{Col}	V_{1y10y}^{Col}	V_{1y20y}^{Col}	V_{1y30y}^{Col}
$2y$	V_{2y2y}^{Col}	V_{2y5y}^{Col}	V_{2y10y}^{Col}	V_{2y20y}^{Col}	V_{2y30y}^{Col}	V_{2y2y}^{Col}	V_{2y5y}^{Col}	V_{2y10y}^{Col}	V_{2y20y}^{Col}	V_{2y30y}^{Col}
$5y$	V_{5y2y}^{Col}	V_{5y5y}^{Col}	V_{5y10y}^{Col}	V_{5y20y}^{Col}	V_{5y30y}^{Col}	V_{5y2y}^{Col}	V_{5y5y}^{Col}	V_{5y10y}^{Col}	V_{5y20y}^{Col}	V_{5y30y}^{Col}
$10y$	V_{10y2y}^{Col}	V_{10y5y}^{Col}	V_{10y10y}^{Col}	V_{10y20y}^{Col}	V_{10y30y}^{Col}	V_{10y2y}^{Col}	V_{10y5y}^{Col}	V_{10y10y}^{Col}	V_{10y20y}^{Col}	V_{10y30y}^{Col}
$15y$	V_{15y2y}^{Col}	V_{15y5y}^{Col}	V_{15y10y}^{Col}	V_{15y20y}^{Col}	V_{15y30y}^{Col}	V_{15y2y}^{Col}	V_{15y5y}^{Col}	V_{15y10y}^{Col}	V_{15y20y}^{Col}	V_{15y30y}^{Col}
$20y$	V_{20y2y}^{Col}	V_{20y5y}^{Col}	V_{20y10y}^{Col}	V_{20y20y}^{Col}	V_{20y30y}^{Col}	V_{20y2y}^{Col}	V_{20y5y}^{Col}	V_{20y10y}^{Col}	V_{20y20y}^{Col}	V_{20y30y}^{Col}

swap annuity, defined as

$$A_p(0) = \sum_{i=1}^{N \times m} \Delta_{i-1} D(0, T_i), \tag{2.2}$$

where N is the swap tenor in number of years, m is the payment frequency per year, Δ_{i-1} is the day count fraction for the period starting at T_{i-1} and ending at T_i, and $D(0, T_i)$ is the discount factor representing the present value at time 0 of 1 unit of currency on T_i. The expectation of the swaption payoff in Equation (2.1) is evaluated under the risk-neutral measure $\mathbb{Q}^*$ associated with the money-market account $B(t) = B(0)e^{\int_0^t r_u du}$. In Swap Market Model, it is more convenient to work under $\mathbb{Q}^A$, the risk-neutral measure associated with the swap annuity as the choice of numeraire (de Jong, Driessen & Pelsser 2001). Using the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}^A}{d\mathbb{Q}^*} = \frac{A_p(T)/A_p(0)}{B(T)/B(0)} = \frac{A_p(T)/A_p(0)}{e^{\int_0^T r_t\, dt}} \tag{2.3}$$

we can apply Girsanov's theorem to change the measure from $\mathbb{Q}^*$ to $\mathbb{Q}^A$:

$$V_p^{\text{Swp}}(0, K) = \mathbb{E}^A\left[\frac{d\mathbb{Q}^*}{d\mathbb{Q}^A} e^{-\int_0^T r_t\, dt} A_p(T)(S(T) - K)^+\right] = A_p(0)\mathbb{E}^A\left[(S_T - K)^+\right], \tag{2.4}$$

where the expectation is now taken under the risk-neutral measure $\mathbb{Q}^A$, associated with the swap annuity numeraire A_p. It should be clear that the forward swap rate is a martingale under the measure $\mathbb{Q}^A$, i.e. $\mathbb{E}^A[S(T)] = S(0)$, which can be calculated directly from the discount curve $D(\cdot, \cdot)$ fitted to the interest rate swap market.

As Pugachevsky (2001) has demonstrated, choosing the appropriate martingale measure to work in is important. Forward swap rates under the risk-neutral measure associated with the swap annuity are used in pricing forward starting swaps and swaptions, although the same rates under the risk-neutral measure associated with the zero-coupon discount bond becomes forward CMS rates to be used to price CMS swaps and CMS caps and floors. In the next section (Section 3), we will show how the unifying model can handle this and price both swaptions and CMS products in the same framework.

Swap-settled swaptions can be priced directly by evaluating the expectation of their payoffs under the risk-neutral measure $\mathbb{Q}^A$, and we can obtain closed-form formulae. To this end, we postulate a stochastic model for the swap rate $S(t)$ in the form of a Swap Market Model (SMM)

$$dS(t) = \sigma S(t) dW^A(t), \tag{2.5}$$

where σ is the lognormal volatility, and $W^A(t) \sim N(0, t)$ is a standard Brownian motion under the risk-neutral measure $\mathbb{Q}^A$, associated with the physical swap annuity numeraire $A_p(0)$. Since this is essentially the Black model for the forward swap rate, the SMM yields closed-form analytical pricing formulae for swap-settled swaptions in the form of

$$\begin{aligned} V_p^{\text{Swp}}(0, K) &= A_p(0)\Big[S(0)\Phi(d_1) - K\Phi(d_2)\Big] \\ V_r^{\text{Swp}}(0, K) &= A_p(0)\Big[K\Phi(-d_2) - S(0)\Phi(-d_1)\Big] \end{aligned} \tag{2.6}$$

where

$$d_1 = \frac{\log\frac{S(0)}{K} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}, \qquad d_2 = d_1 - \sigma\sqrt{T}, \tag{2.7}$$

and $\Phi(\cdot)$ denote the cumulative distribution function for the standard normal distribution.

In the swaption markets, the stochastic alpha-beta-rho (SABR) model proposed by Hagan et al. (2002) is the standard market model used for swaptions pricing. The main advantage of SABR model lies in its ability to express implied volatility as a closed-form analytical formula, allowing swaptions to be priced in a quick and efficient manner. Being able to value swaption portfolio efficiently using analytical formulae is important, as swaptions are used as the basis to price more exotic products, including Bermudan swaptions, callable swaps, spread options, and others. Having an analytical expression for swaption prices significantly speeds up the pricing speed of exotic payoffs. The SABR model postulates that

$$\begin{cases} dS(t) = \alpha(t) S(t)^\beta dW_1^A(t) \\ d\alpha(t) = \nu\alpha(t) dW_2^A(t) \end{cases} \qquad \left\langle dW_1^A(\cdot), dW_2^A(\cdot)\right\rangle(t) = \rho\, dt \tag{2.8}$$

where $W_1^A(t)$ and $W_2^A(t)$ are correlated Brownian motions under the risk-neutral measure $\mathbb{Q}^A$, with a correlation coefficient of ρ. In addition, β is the constant elasticity of variance parameter, also commonly referred to as the "backbone" parameter, $\alpha(t)$ is the stochastic volatility with $\alpha(0) = \alpha$, and ν is the volatility of volatility. A Black model implied volatility can be obtained for this model, which is given by (Hagan et al. 2002):

$$\begin{aligned} \sigma_{\text{SABR}} = & \frac{\alpha}{(S(0)K)^{(1-\beta)/2}\left\{1 + \frac{(1-\beta)^2}{24}\log^2\left(\frac{S(0)}{K}\right) + \frac{(1-\beta)^4}{1920}\log^4\left(\frac{S(0)}{K}\right) + \cdots\right\}} \\ & \times \frac{z}{x(z)} \times \left\{1 + \left[\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(S(0)K)^{1-\beta}} + \frac{1}{4}\frac{\rho\beta\nu\alpha}{(S(0)K)^{(1-\beta)/2}} + \frac{2-3\rho^2}{24}\nu^2\right] T + \cdots\right. \end{aligned} \tag{2.9}$$

where

$$z = \frac{\nu}{\alpha}(S(0)K)^{(1-\beta)/2}\log\left(\frac{S(0)}{K}\right), \tag{2.10}$$

and

$$x(z) = \log\left[\frac{\sqrt{1-2\rho z + z^2} + z - \rho}{1-\rho}\right]. \tag{2.11}$$

Given this implied volatility, we can price swap-settled swaptions by substituting for this volatility in place of σ in the Black formula in Equation (2.6).

The four SABR model parameters, α, β, ν, and ρ, can be calibrated to liquid swaption market quotes by running the following optimization:

$$\min_{\alpha,\,\beta,\,\nu,\,\rho} \sum_{i=1}^{n+k+1} \Big(V^{\text{model}}(0, K_i) - V^{\text{market}}(0, K_i) \Big)^2. \tag{2.12}$$

In practice, it is common for the trading desks to select the β parameter according to the volatility backbone in the swaption market (Zhang & Fabozzi 2016 for further discussion on this issue), and only calibrate α, ν, and ρ to the swaption market, each of which will determine the level, slope, and smile of the implied volatility surface, respectively.

Note that under SMM, swap-settled swaptions pricing can be expressed analytically in the form of a Black formula. This is a major advantage, given the size of swaption portfolio and the use of swaptions as basis for exotic product pricing, it is important to retain analytical tractability for liquid vanilla products traded in large volume.

2.3 Pricing IRR-Settled Swaption

In this section, we will outline the approach widely adopted in the industry to price IRR-settled swaptions. As mentioned in previous sections, two simplifying assumptions are required to arrive at the approximation formula. However, given the efficiency of using the market model along with the SABR implied volatility in pricing swap-settled swaptions, the same approach is used to price IRR-settled swaptions, even though the two simplifying assumptions made are not arbitrage-free. Although the industry on the whole is aware of the issues involved, the error incurred in the approximation formula is often considered to be minor (Brigo & Mercurio 2006).

When market is quoting and trading cash-settled swaptions (denoted here as V^{IRR}), the pricing function becomes, under the forward risk-neutral measure $\mathbb{Q}^T$,

$$V_p^{\text{IRR}}(0, K) = D(0, T)\mathbb{E}^T \left[A_c(S(T))(S(T) - K)^+ \right] \tag{2.13}$$

where A_c is the IRR annuity defined as

$$A_c(S) = \sum_{i=1}^{N\times m} \frac{\frac{1}{m}}{(1+\frac{1}{m}\cdot S)^i}. \tag{2.14}$$

Similar to the swap annuity in Equation (2.2), N is the number of years of the swap tenor and m is the payment frequency. Here $\frac{1}{m}$ is the day count fraction. The expectation is evaluated under the risk-neutral measure $\mathbb{Q}^T$, associated with the discount factor $D(0,T)$ as numeraire, where 0 is the pricing date, and T is the maturity date of the swaption.

It is standard market practice to price IRR-settled swaptions using the following approximation:

$$\begin{aligned} V_p^{\text{IRR}}(0,K) &\overset{(1)}{\approx} D(0,T)A_c(S(0))\mathbb{E}^T\left[(S(T)-K)^+\right] \\ &\overset{(2)}{\approx} D(0,T)A_c(S(0))\mathbb{E}^A\left[(S(T)-K)^+\right], \end{aligned} \tag{2.15}$$

Compared to Equation (2.13), it is obvious that this approximation incurs two errors: (1) the concave (convex) payoff profile of the IRR-settled payer (receiver) swaption is not accounted for when $A_c(S(T))$ is approximated as $A_c(S(0))$, which is known today; and (2) the expectation is taken under the risk-neutral measure $\mathbb{Q}^A$ instead of $\mathbb{Q}^T$.

The reason for the first approximation is to retain analytical tractability. Market prefers to use Black-like formulae for pricing due to its speed and efficiency, along with the intuitions practitioners have built up over time in using them. If one were to keep $A_c(S(T))$ in the expectation, one will no longer be able to obtain a closed-form expression. The reason for the second approximation is to work under the martingale measure of the forward swap rate, even though this incurs an error, as the forward swap rate is not a martingale under the risk-neutral measure $\mathbb{Q}^T$ measure. If we need to evaluate the expectation under $\mathbb{Q}^T$, the value of $S(T)$ will be unknown.

A common justification given for this approximation is as follows: the IRR-settled swaption can also be expressed, via a change of measure, as

$$V_p^{\text{IRR}}(0,K) = A_p(0)\mathbb{E}^A\left[\frac{D(T,T)A_c(T)}{A_p(T)}(S(T)-K)^+\right]. \tag{2.16}$$

If one argues that the ratio of annuities $D(T,T)A_c(T)/A_p(T)$ is slow-varying and can be approximated as its initial value $D(0,T)A_c(0)/A_p(0)$, then one obtains the market approximation formula. There have also been attempts to model this ratio of annuities separately with a simple low-variance process, though it is often difficult to reconcile this simple process for the ratio of annuities with the model for the swap rates.

2.3.1 Constant Maturity Swap (CMS) Payoffs

A CMS payoff pays a swap rate quoted in the market on a given date. The payment can be made either in arrears or in advance, corresponding to the end or the start of each accrual period, respectively. On top of that, a CMS payoff can also be capped or floored. The value of a CMS payoff is sensitive to the shape of the distribution of the swap rate. In other words, the implied volatility smile profile plays an important role in determining the convexity correction of CMS products.

It is well known that one can calculate the CMS rates by convexity adjustment or by static payoff replication (Hagan 2003, Hunt & Kennedy 2000, and Andersen & Piterbarg 2010). Volatility smile has a significant impact, which, in particular, cannot be ignored when pricing CMS swaps or options. Boenkost & Schmidt (2009) demonstrated the importance of evaluating popular exotic interest rate derivatives such as Libor-in-arrears caps or CMS caps by incorporating the volatility smile present in the cap and swaption market. This section covers how CMS products can be priced by static replication to ensure consistency with the volatility smile observed in the IRR-settled swaption market.

Pricing CMS products involve evaluating the payoffs under the risk-neutral measure $\mathbb{Q}^T$. Since this is not the natural martingale measure of the forward swap rate, convexity correction is required to evaluate the expectation of the payoff. Different approaches have been formulated to perform convexity correction for CMS payoffs. In this paper, we focus on the static replication method.

Note that unlike swap-settled swaptions, where the swap annuity A_p is sensitive to the entire yield curve, IRR-settled swaptions can be thought of as contingent claims on the swap rate itself, having an exposure solely to the swap rate S_T, observed at time T. The seminal paper by Breeden & Litzenberger (1978) connected option prices with no-arbitrage state prices, the equivalence of discounted risk-neutral densities. Practical approaches of extracting or inferring the risk-neutral densities using option prices were developed in Bakshi, Kapadia & Madan (2003) and Jiang & Tian (2005). A historical review is provided in Zimmermann (2018). Here we outline the application of this approach to the swaption market.

Consider an IRR-settled payer swaption, which can be valued as

$$V_p^{\text{IRR}}(K) = D(0,T)\int_K^\infty A_c(s)(s-K)^+ f(s)\, ds, \tag{2.17}$$

where $f(s)$ denote the risk-neutral probability density function of the forward swap rate under forward measure $\mathbb{Q}^T$. Differentiating twice with respect to strike K, we obtain an expression for the risk-neutral density

$$f(K) = \frac{\partial^2 V_p^{\text{IRR}}(K)}{\partial K^2} \times \frac{1}{D(0,T)A_c(K)}. \tag{2.18}$$

The same result can also be obtained by differentiating the IRR-settled receiver swaption formula twice. Now suppose we wish to value a CMS contract with the payoff $g(S(T))$ at time T, let $h(K) = g(K)/A_c(K)$, following Carr & Madan (2001), and choosing the rate L as the expansion point, we can show that

$$\begin{aligned} V_{\text{rate}}^{\text{CMS}}(0) &= D(0,T)g(L) + h'(L)[V_p^{\text{IRR}}(L) - V_r^{\text{IRR}}(L)] \\ &\quad + \int_0^L h''(K)V_r^{\text{IRR}}(K)\, dK + \int_L^\infty h''(K)V_p^{\text{IRR}}(K)\, dK \end{aligned} \tag{2.19}$$

where the derivatives of $h(K)$ are given by

$$\begin{aligned} h(K) &= \frac{g(K)}{A_c(K)} \\ h'(K) &= \frac{A_c(K)g'(K) - g(K)A_c'(K)}{A_c(K)^2} \\ h''(K) &= \frac{A_c(K)g''(K) - A_c''(K)g(K) - 2A_c'(K)g'(K)}{A_c(K)^2} + \frac{2A_c'(K)^2 g(K)}{A_c(K)^3}. \end{aligned} \tag{2.20}$$

Choosing the expansion point to be at the forward swap rate, such that $L = S(0)$, a CMS rate payment ($g(S) = S$) for the swap rate at time T can be valued as

$$V_{\text{rate}}^{\text{CMS}}(0) = D(0,T)S(0) + \int_0^{S(0)} h''(K)V_r^{\text{IRR}}(K)dK + \int_{S(0)}^\infty h''(K)V_p^{\text{IRR}}(K)dK \tag{2.21}$$

using static replication, where

$$h''(K) = \frac{-A_c''(K)\cdot K - 2A_c'(K)}{A_c(K)^2} + \frac{2\cdot A_c'(K)^2\cdot K}{A_c(K)^3}. \tag{2.22}$$

Based on the discussions in this section, we see that under SMM, swaptions can be priced via analytical formulae, while CMS products need to be evaluated in a one-dimensional integral. To be specific, the formulae for swap-settled swaptions are exact, but the formulae for IRR-settled swaptions involve two simplifying assumptions and are not arbitrage-free. The evaluation of the CMS static replication integral over the full continuum of payer and receiver IRR-settled swaptions gives rise to the sensitivity to the implied volatility smile profile.

3 Unified Market Model

For the European swaption markets which trade IRR-settled swaptions as the main interest rate volatility instruments, the main limitation in the standard market approach outlined in the previous section stems from formulating the model of the swap rate process by working under $\mathbb{Q}^A$, the risk-neutral measure associated with the swap annuity as numeraire. Jamshidian (1997) has shown that the forward swap rate is the expectation of the CMS rate under the swap annuity measure, while CMS swaps and caps require calculation of CMS under a different, forward measure. From Equation (2.13), we can see that this is also the same measure used for IRR-settled swaption pricing. Instead of working under the risk-neutral measure $\mathbb{Q}^A$ (which is the case for Swap Market Model), we define the unified market model as

$$dS(t) = S(t)(\theta(t)dt + \sigma dW^T(t)), \tag{3.1}$$

where $W^T(t)$ is a standard Brownian motion under $\mathbb{Q}^T$, which is the forward risk-neutral measure associated with the discount bond $D(0,T)$. Note that the swap rate $S(0)$ is not a martingale under $\mathbb{Q}^T$, hence a drift rate of $\theta(t)$ is included in the model, which is a deterministic function of time. Using this model, we want to value the IRR-settled payer swaptions by evaluating the expectation:

$$V_p^{\text{IRR}}(0,K) = D(0,T)\mathbb{E}^T[A_c(S(T))(S(T)-K)^+]. \tag{3.2}$$

The objective is to evaluate the expectations explicitly without making any of the approximations employed in the market approach presented in the previous section, while retaining as much analytical tractability as possible. To this end, we formulate a displaced-diffusion stochastic volatility model under the risk-neutral measure $\mathbb{Q}^T$. The SABR model postulates a constant elasticity of variance (CEV) process for the swap rate. In our formulation, we use the displaced-diffusion process instead

(Rubinstein 1983). The rationale behind this choice is two-fold. First, displaced-diffusion dynamics allow us to retain a larger degree of analytical tractability. Second, negative rates are admissible under the displaced-diffusion process, which is consistent with recent interest rate regime in the euro, Swiss franc, and yen markets. Furthermore, recent research by Svoboda-Greenwood (2009) and Lee & Wang (2012) have investigated the mathematical properties of the displaced-diffusion dynamics and demonstrated its ability to fit the market well.

In the euro market, which is also the largest market for IRR-settled swaptions, the spot swap rates out to approximately $7y$ are negative as of the time of writing in 2020. Hence, for any model to be useful to practitioners, it needs to be able to handle negative forward rates. Although the common displaced-diffusion model formulation

$$dS(t) = \sigma[\beta S(t) + (1-\beta)S(0)]dW(t) \tag{3.3}$$

is able to handle negative rates, there are certain limitation in this approach. First, the lower rate boundary of this displaced-diffusion formulation is $(1-\beta)/\beta S(0)$, and might not be sufficient for the negative rate domain. Second, if the initial forward swap rate $S(0)$ happens to be zero, then the swap rate process actually never moves away from zero. For initial forward swap rates close to but not exactly equal to zero, the formulation may also become numerically unstable. Consequently, we opted for the alternative formulation based on a simple displacement parameter as follows:

$$dS(t) = \sigma\Big[S(t) + \beta\Big]dW(t). \tag{3.4}$$

Let $\sqrt{V(t)}$ denote the stochastic volatility, the model is given by:

$$\begin{cases} dS(t) = \Big[S(t) + \beta\Big]\Big(\theta(t)dt + \sqrt{V(t)}dZ_1^T(t)\Big) \\ dV(t) = \nu V(t) dZ_2^T(t) \end{cases} \tag{3.5}$$

where $Z_1^T(t)$ and $Z_2^T(t)$ are independent Brownian motions under $\mathbb{Q}^T$. Here β is the displaced-diffusion parameter, $\theta(t)$ is the time-dependent drift rate of the swap rate, ν is the volatility of volatility. Note that the swap rate process can be allowed to be correlated with the variance process by writing $dV(t) = \nu V(t)\left(\rho dZ_1^T(t) + \sqrt{1-\rho^2}dZ_2^T(t)\right)$ through the correlation parameter ρ. However, following Hull & White (1987), assuming independence allows one to further simplify the model to obtain semi-analytical pricing formulae, as we elaborate later.

In our model, the stochastic variance is modelled as a lognormal process. We note that the forward swap rate is a martingale under the risk-neutral measure $\mathbb{Q}^A$, so that $\mathbb{E}^A[S(T)] = S(0)$. However, this is not the case under the risk-neutral measure $\mathbb{Q}^T$, so the value of $\mathbb{E}^T[S(T)]$ is unknown, and we need to explicitly calibrate the drift parameter θ in the model. The swap rate process can be solved to yield:

$$S(T) = \Big[S(0) + \beta\Big] \exp\left[\int_0^T \theta(t)\, dt - \frac{1}{2}\int_0^T V(t)\, dt + \int_0^T \sqrt{V(t)}\, dZ_1^T(t)\right] - \beta. \tag{3.6}$$

We define the mean integrated variance over the time period $[0, T]$ as

$$\bar{V} = \frac{1}{T}\int_0^T V(t)\, dt. \tag{3.7}$$

Conditional on this mean integrated variance $\bar{V}$, the term

$$\exp\left[\int_0^T \theta(t)\, dt - \frac{1}{2}\int_0^T V(t)\, dt + \int_0^T \sqrt{V(t)}\, dZ_1^T(t)\right] \tag{3.8}$$

is lognormally distributed, where

$$\int_0^T \theta(t)\, dt - \frac{1}{2}\int_0^T V(t)\, dt + \int_0^T \sqrt{V(t)}\, dZ_1^T(t) \sim N\left(\int_0^T \theta(t)\, dt - \frac{\bar{V}T}{2}, \bar{V}T\right). \tag{3.9}$$

Therefore, we have the following expectation under $\mathbb{Q}^T$:

$$\mathbb{E}^T[S(T)] = \Big[S(0) + \beta\Big] \exp\left(\int_0^T \theta(t)\, dt\right) - \beta. \tag{3.10}$$

This is a direct consequence of our explicit choice not to work under the natural martingale measure of the forward swap rate, and the term $\int_0^T \theta(t)\, dt$ accounts for the drift. Note that Equation (3.10) evaluates to $S(0)$ if $\forall t \in [0, T] : \ \theta(t) = 0$. It should be clear that the distribution of the forward swap rate $S(T)$ depends on the paths followed by the stochastic variance process.

3.1 Pricing IRR-Settled Swaptions

Theorem 3.1 (IRR-Settled Swaptions Pricing Formula). *Under the unified market model formulated in Equation* (3.5)*, and the conditional forward swap rate distribution in Equation* (3.9)*, IRR-settled swaptions* *can be priced using the following formula:*

$$
\begin{aligned}
V^{IRR}(0,K) &= \int_0^\infty \boldsymbol{DD}(S(0),K,\bar{V},T)\ \psi(\bar{V})\ d\bar{V} \\
&\approx \boldsymbol{DD}(S(0),K,\bar{V},T) + \frac{1}{2}\frac{\partial^2 \boldsymbol{DD}}{\partial \bar{V}^2}\Big(\mathbb{E}[\bar{V}^2] - \mathbb{E}[\bar{V}]^2\Big) \\
&\quad + \frac{1}{6}\frac{\partial^3 \boldsymbol{DD}}{\partial \bar{V}^3}\Big(\mathbb{E}[\bar{V}^3] - 3\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}^2] + 3\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^2 - \mathbb{E}[\bar{V}]^3\Big) \\
&\quad + \frac{1}{24}\frac{\partial^4 \boldsymbol{DD}}{\partial \bar{V}^4}\Big(\mathbb{E}[\bar{V}^4] - 4\mathbb{E}[\bar{V}^3]\mathbb{E}[\bar{V}] + 6\mathbb{E}[\bar{V}^2]\mathbb{E}[\bar{V}]^2 - 4\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^3 + \mathbb{E}[\bar{V}]^4\Big) \\
&\quad + \frac{1}{120}\frac{\partial^5 \boldsymbol{DD}}{\partial \bar{V}^5}\Big(\mathbb{E}[\bar{V}^5] - 5\mathbb{E}[\bar{V}^4]\mathbb{E}[\bar{V}] + 10\mathbb{E}[\bar{V}^3]\mathbb{E}[\bar{V}]^2 - 10\mathbb{E}[\bar{V}^2]\mathbb{E}[\bar{V}]^3 + 5\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^4 - \mathbb{E}[\bar{V}]^5\Big) + \cdots
\end{aligned}
\tag{3.11}
$$

where $DD(\cdot)$ *denote the one-dimensional integral*

$$
\boldsymbol{DD}\Big(S(0),K,\bar{V},T\Big) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{S(T)}\left[1 - \frac{\frac{1}{m}}{\big(1+\frac{1}{m}\cdot S(T)\big)^{N\times m}}\right]\Big(S(T)-K\Big)^{+} e^{-\frac{x^2}{2}}\ dx,
\tag{3.12}
$$

for payer swaptions, and

$$
\boldsymbol{DD}\Big(S(0),K,\bar{V},T\Big) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{S(T)}\left[1 - \frac{\frac{1}{m}}{\big(1+\frac{1}{m}\cdot S(T)\big)^{N\times m}}\right]\Big(K-S(T)\Big)^{+} e^{-\frac{x^2}{2}}\ dx,
\tag{3.13}
$$

for receiver swaptions. The moments of $\bar{V}$ *are given by*

$$
\begin{aligned}
\mathbb{E}\left[\bar{V}\right] &= V_0, \\
\mathbb{E}\left[\bar{V}^2\right] &= \frac{V_0^2}{T^2}\frac{2\left(e^{\nu^2 T} - \nu^2 T - 1\right)}{\nu^4}, \\
\mathbb{E}\left[\bar{V}^3\right] &= \frac{V_0^3}{T^3}\frac{e^{3\nu^2 T} - 9e^{\nu^2 T} + 6\nu^2 T + 8}{3\nu^6}, \\
\mathbb{E}\left[\bar{V}^4\right] &= \frac{V_0^4}{T^4}\frac{2\left(e^{6\nu^2 T} + 54e^{\nu^2 T} - 10e^{3\nu^2 T} - 30\nu^2 T - 45\right)}{45\nu^8}, \\
\mathbb{E}\left[\bar{V}^5\right] &= \frac{V_0^5}{T^5}\frac{200e^{3\nu^2 T} - 35e^{6\nu^2 T} + 3e^{10\nu^2 T} - 840e^{\nu^2 T} + 84(8+5\nu^2 T)}{630\nu^{10}}.
\end{aligned}
\tag{3.14}
$$

Proof. To price IRR-settled swaptions, we use the unified market model and work under the risk-neutral measure $\mathbb{Q}^T$

$$
\frac{V^{\text{IRR}}(0,K)}{D(0,T)} = \mathbb{E}^T\left[\frac{V^{\text{IRR}}(T,K)}{D(T,T)}\right].
\tag{3.15}
$$

We start by writing

$$V(0,K) = D(0,T)\int_0^\infty \int_0^\infty A_c(S(T))(S(T)-K)^+\eta(S(T),\bar{V})\; dS(T)\; d\bar{V}. \tag{3.16}$$

Here $\eta(S(T),\bar{V})$ is the joint probability density function of the forward swap rate and the mean integrated variance processes. Following our assumption that the forward swap rate process is uncorrelated with the variance process, we can write $\eta(S(T),\bar{V}) = \psi(\bar{V})\xi(S(T)|\bar{V})$, where ψ is the probability density function for the mean integrated variance $\bar{V}$, while ξ is the conditional probability density function for the swap rate $S(T)$ given the mean integrated variance, which is assumed to be independent of the swap rate process. Hence we have

$$V(0,K) = D(0,T)\int_0^\infty \left(\int_0^\infty A_c(S(T))(S(T)-K)^+\xi(S(T)|\bar{V})\; dS(T)\right)\psi(\bar{V})\; d\bar{V}. \tag{3.17}$$

Conditional on the integrated variance $\bar{V}$, the swap rate process $S(T)$ is shifted-lognormally distributed, and the inner integral for a payer swaption can therefore be written as

$$\mathrm{DD}\Big(S(0),K,\bar{V},T\Big) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{S(T)}\left[1-\frac{\frac{1}{m}}{\left(1+\frac{1}{m}\cdot S(T)\right)^{N\times m}}\right]\Big(S(T)-K\Big)^+ e^{-\frac{x^2}{2}}\; dx, \tag{3.18}$$

where, conditional on $\bar{V}$, the distribution of the forward swap rate $S(T)$ is given by equations (3.6)-(3.9). This is an one-dimensional integral that can be evaluated numerically. Therefore, the value of an IRR-settled swaption is given by

$$V(0,K) = D(0,T)\int_0^\infty \mathrm{DD}\left(S(0),K,\bar{V},T\right)\; \psi\left(\bar{V}\right)\; d\bar{V}, \tag{3.19}$$

Given the stochastic process of the variance, we can obtain the expression for $\bar{V}$ as

$$\bar{V} = V_0 + \frac{1}{T}\int_0^T \nu V_u(T-u)\; dW_u. \tag{3.20}$$

Note that the distribution of the mean integrated variance $\bar{V}$ as given in Equation (3.7) is unknown. Nevertheless, in the stochastic volatility model formulated by Hull & White (1987), they point out that although it is impossible to obtain an analytic form for the distribution of $\bar{V}$, its moments $M_X(j) = \mathbb{E}\left[\bar{V}^j\right]$ can be derived analytically. In their paper, they provided the first three moments of the mean integrated variance. In this paper, we also derive and provide the fourth and fifth moments to make up a total of five moments.

Under this formulation, the swaption price becomes the weighted sum over the displaced-diffusion formula $DD(\cdot)$ for different integrated variance. This intuitive and elegant result is often referred to as the "mixing" theorem, and is first derived by Hull & White (1987). Finally, by Taylor expansion, we write the value of the IRR-settled payer swaption as follows:

$$
\begin{aligned}
&\int_0^\infty \mathrm{DD}(S(0), K, \bar{V}, T)\ \psi(\bar{V})\ d\bar{V} \\
&\quad \approx \mathrm{DD}(S(0), K, \bar{V}, T) + \frac{1}{2}\frac{\partial^2 \mathrm{DD}}{\partial \bar{V}^2}\Big(\mathbb{E}[\bar{V}^2] - \mathbb{E}[\bar{V}]^2\Big) \\
&\qquad + \frac{1}{6}\frac{\partial^3 \mathrm{DD}}{\partial \bar{V}^3}\Big(\mathbb{E}[\bar{V}^3] - 3\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}^2] + 3\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^2 - \mathbb{E}[\bar{V}]^3\Big) \\
&\qquad + \frac{1}{24}\frac{\partial^4 \mathrm{DD}}{\partial \bar{V}^4}\Big(\mathbb{E}[\bar{V}^4] - 4\mathbb{E}[\bar{V}^3]\mathbb{E}[\bar{V}] + 6\mathbb{E}[\bar{V}^2]\mathbb{E}[\bar{V}]^2 - 4\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^3 + \mathbb{E}[\bar{V}]^4\Big) \\
&\qquad + \frac{1}{120}\frac{\partial^5 \mathrm{DD}}{\partial \bar{V}^5}\Big(\mathbb{E}[\bar{V}^5] - 5\mathbb{E}[\bar{V}^4]\mathbb{E}[\bar{V}] + 10\mathbb{E}[\bar{V}^3]\mathbb{E}[\bar{V}]^2 - 10\mathbb{E}[\bar{V}^2]\mathbb{E}[\bar{V}]^3 + 5\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^4 - \mathbb{E}[\bar{V}]^5\Big) + \cdots
\end{aligned}
\tag{3.21}
$$

This results in an IRR-settled swaption pricing formula involving one-dimensional numerical integrals in the unified market model.

■

3.2 Pricing Swap-Settled Swaptions

Theorem 3.2 (Swap-Settled Swaptions Pricing Formula). *Under the unified market model formulated in Equation* (3.5)*, and the conditional forward swap rate distribution in Equation* (3.9)*, swap-settled swaptions can be priced using the following formula:*

$$
\begin{aligned}
V(0,K) &= \int_0^\infty \tilde{DD}(S(0), K, \bar{V}, T)\ \psi(\bar{V})\ d\bar{V} \\
&\approx \tilde{DD}(S(0), K, \bar{V}, T) + \frac{1}{2}\frac{\partial^2 \tilde{DD}}{\partial \bar{V}^2}\Big(\mathbb{E}[\bar{V}^2] - \mathbb{E}[\bar{V}]^2\Big) \\
&\quad + \frac{1}{6}\frac{\partial^3 \tilde{DD}}{\partial \bar{V}^3}\Big(\mathbb{E}[\bar{V}^3] - 3\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}^2] + 3\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^2 - \mathbb{E}[\bar{V}]^3\Big) \\
&\quad + \frac{1}{24}\frac{\partial^4 \tilde{DD}}{\partial \bar{V}^4}\Big(\mathbb{E}[\bar{V}^4] - 4\mathbb{E}[\bar{V}^3]\mathbb{E}[\bar{V}] + 6\mathbb{E}[\bar{V}^2]\mathbb{E}[\bar{V}]^2 - 4\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^3 + \mathbb{E}[\bar{V}]^4\Big) \\
&\quad + \frac{1}{120}\frac{\partial^5 \tilde{DD}}{\partial \bar{V}^5}\Big(\mathbb{E}[\bar{V}^5] - 5\mathbb{E}[\bar{V}^4]\mathbb{E}[\bar{V}] + 10\mathbb{E}[\bar{V}^3]\mathbb{E}[\bar{V}]^2 - 10\mathbb{E}[\bar{V}^2]\mathbb{E}[\bar{V}]^3 + 5\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^4 - \mathbb{E}[\bar{V}]^5\Big) + \cdots
\end{aligned}
\tag{3.22}
$$

where

$$\tilde{DD}\Big(S(0), K, \bar{V}, T\Big) = A_p(t)\left[S'(0)\Phi\left(\frac{\log\frac{S'(0)}{K'} + \frac{\bar{V}'T}{2}}{\sqrt{\bar{V}'T}}\right) - K'\Phi\left(\frac{\log\frac{S'(0)}{K'} - \frac{\bar{V}'T}{2}}{\sqrt{\bar{V}'T}}\right)\right], \tag{3.23}$$

for payer swaptions, and

$$\tilde{DD}\Big(S(0), K, \bar{V}, T\Big) = A_p(t)\left[S'(0)\Phi\left(\frac{\log\frac{S'(0)}{K'} + \frac{\bar{V}'T}{2}}{\sqrt{\bar{V}'T}}\right) - K'\Phi\left(\frac{\log\frac{S'(0)}{K'} - \frac{\bar{V}'T}{2}}{\sqrt{\bar{V}'T}}\right)\right], \tag{3.24}$$

for receiver swaptions, and

$$K' = K + \beta, \qquad S'(0) = S(0) + \beta, \qquad \textit{and} \qquad \bar{V}' = \sqrt{\beta}\bar{V}. \tag{3.25}$$

Proof. To price swap-settled swaptions, we need to change the measure from $\mathbb{Q}^T$ to $\mathbb{Q}^A$, which involves changing the choice of numeraire from $D(t,T)$ to $A_p(t)$. We note that $\mathbb{Q}^A$ is a risk-neutral measure equivalent to $\mathbb{Q}^T$ under the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}^A}{d\mathbb{Q}^T} = \exp\left[-\int_0^t \kappa(u)\, dZ(u) - \frac{1}{2}\int_0^t \kappa(u)^2\, du\right]. \tag{3.26}$$

Using Girsanov's theorem, given that Z_1 and Z_2 are independent, we have

$$\begin{aligned} Z_1^A(t) &= Z_1^T(t) + \int_0^t \kappa(u)\, du \\ Z_2^A(t) &= Z_2^T(t) \end{aligned} \tag{3.27}$$

and after substituting we obtain the following model for swap-settled swaptions:

$$\begin{cases} dS(t) = \Big[S(t) + \beta\Big]\left[\Big(\theta(t) - \kappa(t)\sqrt{V(t)}\Big)\, dt + \sqrt{V(t)} dZ_1^A(t)\right] \\ dV(t) = \nu V(t) dZ_2^A(t) \end{cases}. \tag{3.28}$$

However, we know that $\mathbb{E}^A[S(T)] = S(0)$, since we are now working under the natural martingale measure $\mathbb{Q}^A$ of the swap rate process, the model simplifies to

$$\begin{cases} dS(t) = \sqrt{V(t)}\Big[S(t) + \beta\Big] dZ_1^A(t) \\ dV(t) = \nu V(t) dZ_2^A(t) \end{cases}. \tag{3.29}$$

For swap-settled swaptions, we work under the martingale measure $\mathbb{Q}^A$ with its associated numeraire A_p and write:

$$\frac{V^{\text{Swp}}(0,K)}{A_p(0)} = \mathbb{E}^A\left[\frac{V^{\text{Swp}}(T,K)}{A_p(T)}\right], \tag{3.30}$$

which yields (for payer swaptions)

$$V_p^{\text{Swp}}(0,K) = A_p(0)\int_0^\infty \int_0^\infty (S(T)-K)^+ \eta(S(T),\bar{V})\, dS(T)\, d\bar{V}. \tag{3.31}$$

As before, under our assumption that the forward swap rate process is independent from the variance process, we have $\eta(S(T),\bar{V}) = \psi(\bar{V})\xi(S(T)|\bar{V})$, and

$$V_p^{\text{Swp}}(0,K) = \int_0^\infty \left(A_p(0)\int_0^\infty (S(T)-K)^+ \xi(S(T)|\bar{V})\, dS(T)\right) \psi(\bar{V})\, d\bar{V}. \tag{3.32}$$

Conditional on the integrated variance $\bar{V}$, the displaced-diffusion price in the inner integral can therefore be written as

$$\tilde{\text{DD}}\Big(S(0),K,\bar{V},T\Big) = A_p(t)\left[S'(0)\Phi\left(\frac{\log\frac{S'(0)}{K'} + \frac{\bar{V}'T}{2}}{\sqrt{\bar{V}'T}}\right) - K'\Phi\left(\frac{\log\frac{S'(0)}{K'} - \frac{\bar{V}'T}{2}}{\sqrt{\bar{V}'T}}\right)\right], \tag{3.33}$$

where

$$K' = K+\beta, \qquad S'(0) = S(0)+\beta, \qquad \text{and} \qquad \bar{V}' = \sqrt{\beta}\bar{V}. \tag{3.34}$$

Therefore, the value of an IRR-settled payer swaption is given by

$$V_p(0,K) = \int_0^\infty \tilde{\text{DD}}\left(S(0),K,\bar{V},T\right)\ \psi\left(\bar{V}\right)\ d\bar{V}, \tag{3.35}$$

where ψ is the probability density function of the mean integrated variance $\bar{V}$. Finally, by Taylor expansion, we write the value of the swap-settled swaption as follows:

$$\begin{aligned}
&\int_0^\infty \tilde{\text{DD}}(S(0),K,\bar{V},T)\, \psi(\bar{V})\, d\bar{V} \\
&\quad\approx \tilde{\text{DD}}(S(0),K,\bar{V},T) + \frac{1}{2}\frac{\partial^2\tilde{\text{DD}}}{\partial\bar{V}^2}\Big(\mathbb{E}[\bar{V}^2] - \mathbb{E}[\bar{V}]^2\Big) \\
&\qquad + \frac{1}{6}\frac{\partial^3\tilde{\text{DD}}}{\partial\bar{V}^3}\Big(\mathbb{E}[\bar{V}^3] - 3\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}^2] + 3\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^2 - \mathbb{E}[\bar{V}]^3\Big) \\
&\qquad + \frac{1}{24}\frac{\partial^4\tilde{\text{DD}}}{\partial\bar{V}^4}\Big(\mathbb{E}[\bar{V}^4] - 4\mathbb{E}[\bar{V}^3]\mathbb{E}[\bar{V}] + 6\mathbb{E}[\bar{V}^2]\mathbb{E}[\bar{V}]^2 - 4\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^3 + \mathbb{E}[\bar{V}]^4\Big) \\
&\qquad + \frac{1}{120}\frac{\partial^5\tilde{\text{DD}}}{\partial\bar{V}^5}\Big(\mathbb{E}[\bar{V}^5] - 5\mathbb{E}[\bar{V}^4]\mathbb{E}[\bar{V}] + 10\mathbb{E}[\bar{V}^3]\mathbb{E}[\bar{V}]^2 - 10\mathbb{E}[\bar{V}^2]\mathbb{E}[\bar{V}]^3 + 5\mathbb{E}[\bar{V}]\mathbb{E}[\bar{V}]^4 - \mathbb{E}[\bar{V}]^5\Big) + \cdots
\end{aligned} \tag{3.36}$$

and the moments of $\bar{V}$ are the same as provided in Equations (3.14). ■

This results in a swap-settled swaption pricing formula in closed-form due to the Black-like formula. Furthermore, since our model is formulated as a unified market model under $\mathbb{Q}^T$, CMS products can be priced readily. Thus, pricing formulae for CMS caplet and floorlet can also be expressed as Black-like formulae based on the same approach as swaptions in the unified framework.

4 Results and Analyses

This section presents the analyses based on our unified market model on a range of topics that the approximation formula failed to address. We begin by calibrating the unified model to liquid market swaption quotes, and show that it can match market instruments as close as existing models. We move on to demonstrate that if we do not make the same simplifying assumptions required to arrive at the approximation formula, then IRR-settled payer and receiver swaptions cannot be priced with the same implied volatility surface. Analyses are also performed to quantify the error incurred by making the two simplifying assumptions to arrive at the approximation formula. Next, we explore the pricing of zero-wide collars and ITM swaptions. Finally, we investigate the sensitivities of our model parameters, and demonstrate the existence of negative vega.

4.1 Model Calibration

As we have elaborated in previous sections, although it is difficult to obtain an analytic form for $\bar{V}$, the independence assumption (between the forward swap rate and the variance processes) allows one obtain efficient pricing formulae for both IRR-settled and swap-settled swaptions through the use of a Taylor expansion. Similarly, we define the integrated drift term $\bar{\theta}$ as

$$\bar{\theta} = \int_0^T \theta(t)\, dt. \tag{4.1}$$

There are consequently 4 model parameters: $\bar{\theta}$, V_0, β, and ν to be calibrated for each expiry-tenor swaption chain. In order to fit the entire swaption volatility cube, we will need a set of parameters for each of the expiry-tenor pair. To calibrate our displaced-diffusion stochastic volatility model, we run the following optimization

$$\min_{\substack{\bar{\theta},\, V_0 \\ \nu,\, \beta}} \sum_{i=1}^{n} \Big(V^{\text{model}}(0, K_i) - V^{\text{market}}(0, K_i) \Big)^2, \tag{4.2}$$

where n is the number of swaptions available for a given expiry-tenor pair.

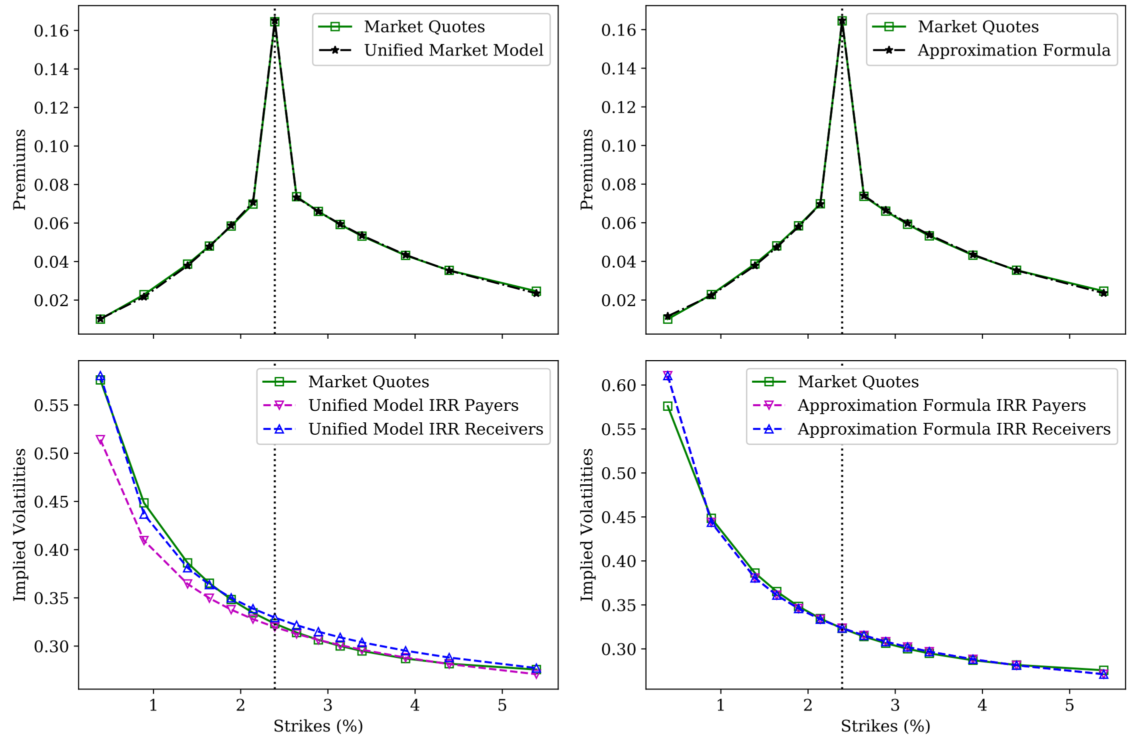
The market convention is to quote and trade at-the-money (ATM) straddles and out-of-the-money (OTM) payer and receiver swaptions. If the pricing function is consistent and arbitrage-free, then put-call parity will hold for European exercise type, and one can infer ITM swaption prices from OTM prices using the same implied volatility of the same strike. It will become clear in this section that this approach only holds for swap-settled swaptions, but not for IRR-settled swaptions. A unified model should be able to, once calibrated, extrapolate market quotes and produce two different implied volatility smiles for IRR-settled payer and receiver swaptions. In addition, it should also be able to produce a third distinct smile for swap-settled swaptions.

Exhibit II plots the market premiums and the quoted implied volatilities of the IRR-settled swaptions for a 10y10y swaption chain, overlaid with the unified market model's premiums and implied volatilities calculated using a calibrated model. Following market's quoting convention, all rates and strikes have been shifted up by 3%. Using standard market convention, implied volatility (σ_{im}) is calculated as

$$V^{\text{IRR}}(0, K) = D(0, T) A_c(0) \text{Black}(S(0), K, \sigma_{im}, T) \tag{4.3}$$

Note that this is essentially the approximation formula, with σ_{im} being the SABR implied volatility.

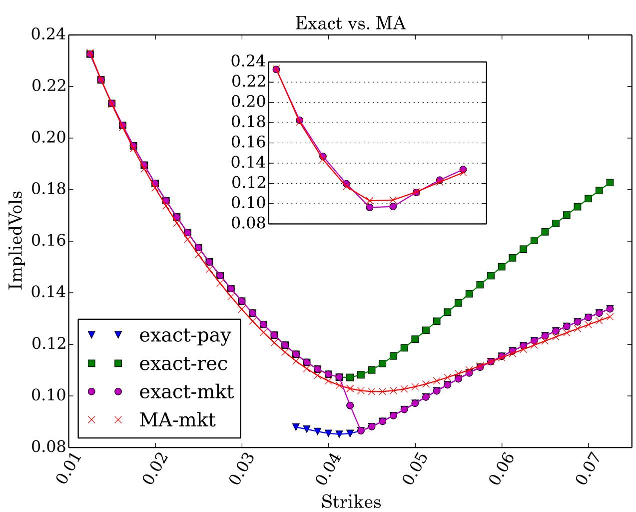
Exhibit II: Comparison of premiums and implied volatilities market quotes *vs* model based values.



To provide a basis for comparison, we also include the model price calculated using the approximation formula with a calibrated SABR implied volatility. In the upper figures, it should be clear that our unified market model prices are able to match observed market premiums extremely well once calibrated (upper left), comparable to the widely used approximation formula (upper right). In the lower figures, it is immediately apparent that the IRR-settled payer and receiver swaptions have different implied volatilities under the unified model (lower left), in stark contrast to the approximation formula, where the payers and receivers have identical implied volatilities (lower right). As discussed in the earlier sections, the implied volatilities of payers and receivers only become equivalent when the Black model can be used to price them consistently. This is the case for swap-settled swaptions, but for IRR-settled swaptions, two simplifying assumptions are required to arrive at the Black model as an approximation formula. These two simplifying assumptions, which separate the IRR annuity from the payoff profile, and use the incorrect risk-neutral measure so that $S(t)$ is a martingale, results in IRR-settled payer and receiver swaptions implied volatilities being indistinguishable.

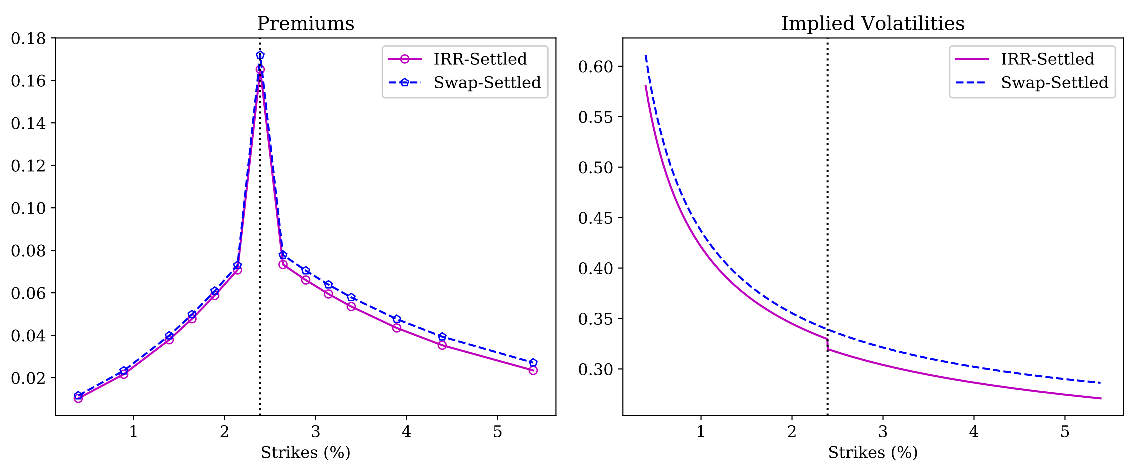
As we have documented in Exhibit I, in the market only OTM and ATM implied volatilities are quoted, and the strikes are spaced at intervals in the range of 25, 50, or 100 bps. Our preceding analysis has highlighted the fact that IRR-settled payer and receiver swaptions ought to be considered separately when considering their implied volatilities. Given the rather coarse strike spacing in market quotes, the kink in the implied volatility due to the transition from OTM receivers to OTM payers is not directly visible. This sudden switch of implied volatilities between the payers and receivers only manifest when then strike resolution is refined. We illustrate this effect using our unified pricing model. Exhibit III compares the implied volatilities calculated using our calibrated exact unified model (labeled as "exact-pay" for payers, "exact-rec" for receivers, and "exact-mkt" for the market OTM/ATM quotes) *vs.* the calibrated approximation formula (labeled as "MA-mkt"). As expected, under the unified model, the implied volatility curve ("exact-mkt") shows an kink across the ATM strike. In addition, the implied volatility curves for payers and receivers are different. On the other hand, no kink in the implied volatilities is observed for the approximation formula ("MA-mkt"). We emphasize that this abrupt jump is only observable when the strike grid is sufficiently refined. As depicted in the inset of Exhibit III, under the standard strike spacing used in market quotes, we will fail to perceive this as a kink, and get the impression that both approaches give rise to very comparable implied volatilities.

Exhibit III: Jump in implied volatility curve.



Next, we move on to illustrate how our unified model is able to price swap-settled and IRR-settled swaptions in a consistent manner. Exhibit IV compares the premiums and implied volatilities of IRR-settled and swap-settled swaptions, both priced using the unified model. We see that our unified model is able to clearly distinguish the difference in settlement methods, and that the kink in implied volatility curve only manifests for the IRR-settled swaptions. This is expected, as unlike IRR-settled swaptions, the implied volatilities for swap-settled payer and receiver swaptions are equivalent. A quick inspection of Equations (3.18) and (3.33) also shows that under the unified model, IRR-settled swaptions are priced with a 1-D integral of the convex/concave payoff function under the forward risk-neutral measure, while swap-settled swaptions are priced with a Black-like formula under the swap annuity measure. On the other hand, if one were to use the approximation formula, swap-settled and IRR-settled swaptions will share the same implied volatilities, and the difference in prices will only stem from the different annuities, which clearly misses out the distinctive differences between the two settlement types.

Exhibit IV: Pricing swaptions using the approximation formula and the unified market model.



Market is accustomed to use the same implied volatility surface to price both payer and receiver swaptions. If this is true, then we can price ITM swaptions using the same implied volatility calibrated to OTM swaptions. For swap-settled swaptions, we have:

$$\begin{aligned} V_p^{\text{Swp}}(0,K) - V_r^{\text{Swp}}(0,K) &= A_p(0)[S(0)(\Phi(d_1)+\Phi(-d_1)) - K(\Phi(d_2)+\Phi(-d_2))] \\ &= A_p(0)(S(0)-K). \end{aligned} \tag{4.4}$$

If the d_1 and d_2 used in the payer and receiver swaptions are identical, which occurs when the same implied volatility is used for both payer and receiver at the same strike, then this relationship is always satisfied. This is valid for swap-settled swaptions only, but not for IRR-settled swaptions.

Under the unified framework, it should be clear that if we were to use a Black-like formula for the IRR-settled swaptions, then the implied volatility for payer and receiver must be different. To see this, simply note that when $K = S(0)$, the Black formula will yield a zero premium for zero-wide collars. For these to be worth a non-zero amount, we need

$$\begin{aligned} V_p^{\text{IRR}}(0,K) - V_r^{\text{IRR}}(0,K) &= D(0,T)\mathbb{E}^T[A_c(T)(S(T)-K)] \\ &= D(0,T)A_c(0)\Big[S(0)(\Phi(d_1(\sigma_p)) + \Phi(-d_1(\sigma_r))) - K(\Phi(d_2(\sigma_p)) + \Phi(-d_2(\sigma_r)))\Big] \end{aligned} \tag{4.5}$$

where σ_p and σ_r are the implied volatilities for the IRR-settled payer and receiver swaptions, respectively. This goes to show that IRR-settled payer and receiver swaptions of the same strike cannot share the same implied volatility, otherwise zero-wide collars will always be worth zero. Any term structure or short rate model that exactly captures the payoff of these zero-wide collars will also show that these instruments have a non-zero price.

4.2 Decomposition of Errors Incurred in the Approximation Formula

In this section we use our unified model to investigate the errors incurred in the approximation formula when making the two critical simplifying assumptions. To this end, we study the price sensitivities when we apply the same simplifying assumptions to our model. The errors incurred for payers will be:

$$\begin{aligned} V_p^{\text{IRR}}(0,K) &\overset{(a)}{=} D(0,T)\mathbb{E}^T[A_c(S(T))(S(T)-K)^+], \\ &\overset{(b)}{\approx} D(0,T)A_c(S(0))\mathbb{E}^T[(S(T)-K)^+] \\ &\overset{(c)}{\approx} D(0,T)\mathbb{E}^A[A_c(S(T))(S(T)-K)^+] \\ &\overset{(d)}{\approx} D(0,T)A_c(S(0))\mathbb{E}^A[(S(T)-K)^+], \end{aligned} \tag{4.6}$$

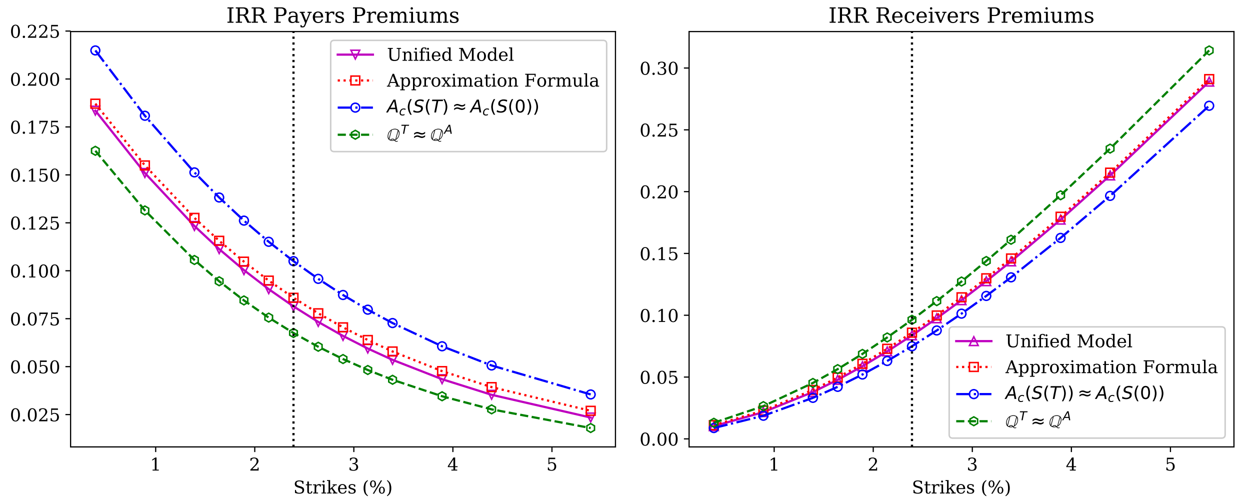
while the errors incurred for receivers will be:

$$
\begin{aligned}
V_r^{\mathrm{IRR}}(0,K) &\overset{(a)}{=} D(0,T)\mathbb{E}^T[A_c(S(T))(K-S(T))^+], \\
&\overset{(b)}{\approx} D(0,T)A_c(S(0))\mathbb{E}^T[(K-S(T))^+] \\
&\overset{(c)}{\approx} D(0,T)\mathbb{E}^A[A_c(S(T))(K-S(T))^+] \\
&\overset{(d)}{\approx} D(0,T)A_c(S(0))\mathbb{E}^A[(K-S(T))^+].
\end{aligned}
\tag{4.7}
$$

In the derivation above, (a) labels the exact formulation of our unified model, (b) labels the simplifying assumption of ignoring the concave/convex payoff profiles of IRR-settlement, (c) labels the simplifying assumption of switching to the swap annuity measure to evaluate the expectations, and (d) labels the approximation formula of applying both simplifying assumptions simultaneously.

Exhibit V plots the IRR-settled payer and receiver swaption premiums calculated using (a) the unified model (triangles), (b) the simplified model which ignores the concave/convex payoffs due to the IRR-annuity (circles), (c) the simplified model which uses the swap-annuity measure (hexagons), and (d) the final approximation formula (squares). It is interesting to note that although both simplification (b) and (c) result in deviation from the right price (a), they err on the opposite side, thereby offsetting and bringing the approximation formula (d) to be closer to the right price than solely making either one of the two simplification. For IRR-settled payer swaptions, simplification (b) inflates the swaption prices, as having the concave IRR-annuity function out of the expectation reduce the discounting effects, while simplification (c) deflates them, since evaluating the expectation under the swap measure underestimate the forward swap rate, thereby reducing the swaption price. The same argument can be also applied to IRR-settled receiver swaptions—in this case, ignoring the effect of the IRR-annuity for receivers leads to underpricing, while using the swap measure underestimates the swap rates, inflating the receiver swaption prices.

Exhibit V: Errors incurred in the two simplifying assumptions in the approximation formula.

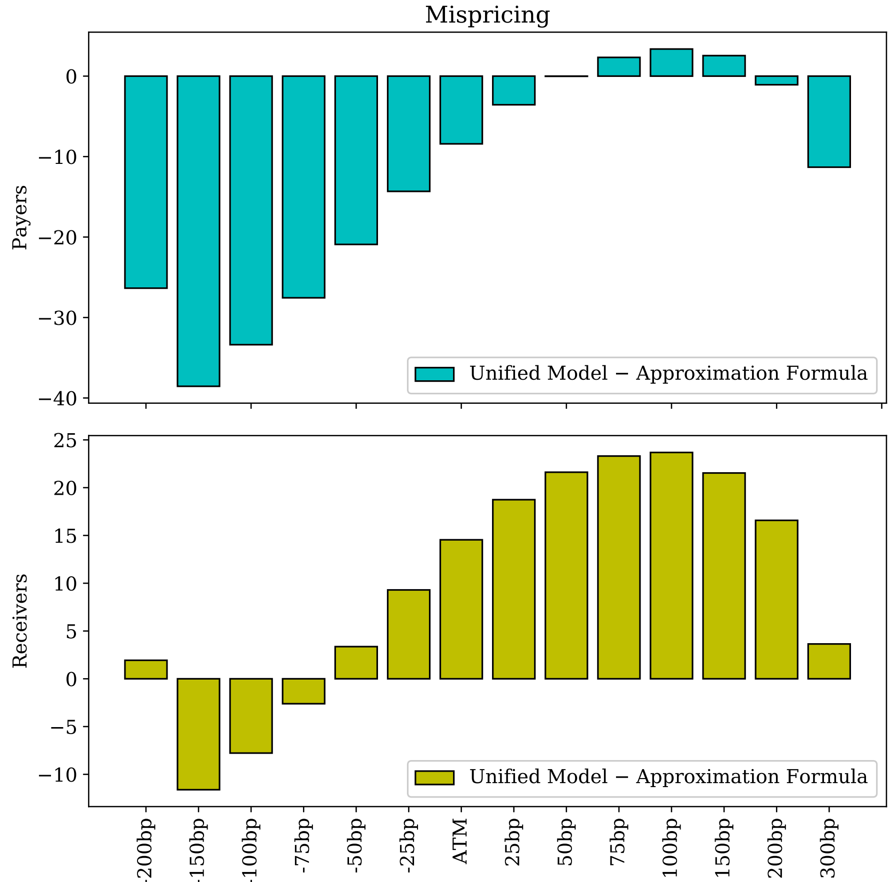


Our analysis presented here provides an explanation to why market participants have been able to use the approximation formula without incurring significant discrepancies—individually the two simplifying assumptions incur sizeable error, but due to their offsetting nature, the overall error incurred is dramatically smaller, and much closer to the accurate price.

4.3 ITM IRR-settled Swaptions & Zero-wide Collars

An immediate consequent of having different implied volatility curves for payer and receiver IRR-settled swaptions is that ITM swaptions could be mispriced due to the lack of calibration data. It is important to note that although the approximation formula is capable of fitting OTM swaptions and ATM straddle prices very well, when applying it to the valuation of ITM swaptions, a large mispricing could manifest given that it does not capture the payoff profile accurately. Exhibit VI shows the mispricing of ITM IRR-settled swaptions, defined as the difference between our unifying model and the approximation formula. In general, the approximation formula overprices low strikes ITM payers (upper figure) and underprices high strikes ITM receivers (lower figure). This observation has important consequences on the risk management of a swaption portfolio. For instance, as the market moves into low rates regime after significant rate cuts, a large number of receiver swaptions also move into-the-money.

Exhibit VI: Mispricing of in-the-money swaptions.



As we have discussed in previous sections, using the approximation formula, IRR-settled payer and receiver swaptions are priced as follows:

$$\begin{cases} V_p^{\text{IRR}}(0,K) \approx D(0,T)A_c(S(0))\mathbb{E}^A[(S(T)-K)^+] \\ V_r^{\text{IRR}}(0,K) \approx D(0,T)A_c(S(0))\mathbb{E}^A[(K-S(T))^+]. \end{cases} \tag{4.8}$$

This implies the following put-call parity relationship:

$$V_p^{\text{IRR}}(0,K) - V_r^{\text{IRR}}(0,K) \approx D(0,T)A_c(S(0))(S(0)-K). \tag{4.9}$$

When $K = \mathbb{E}^A[S(T)] = S(0)$, i.e. when the payer and receiver swaptions are both struck at the forward swap rate, we have $V_p^{\text{IRR}}(0,K) \approx V_r^{\text{IRR}}(0,K)$, and consequently ATM zero-wide collars (struck at the forward swap rate) will be worth 0 under the approximation formula. This is clearly incorrect, and is a direct consequence of the simplifying assumptions made in the standard approach. In fact, euro and sterling markets have begun quoting non-zero premiums for these zero-wide collars (Lutz 2015 and Pietersz & Sengers 2017). The unified model presented in this paper will allow one to price zero-wide collars in a consistent manner once calibrated to the swaption market. Note that

$$V_p^{\text{IRR}}(0,K) - V_r^{\text{IRR}}(0,K) = D(0,T)\mathbb{E}^T[A_c(S(T))(S(T)-K)]. \tag{4.10}$$

Since in the unified model, we do not work under the martingale measure of the swap rate, zero-wide collar will be priced consistently when struck at the forward swap rate. In fact, it can be shown that we only have $V_p^{\text{IRR}}(0,K) = V_r^{\text{IRR}}(0,K)$ when

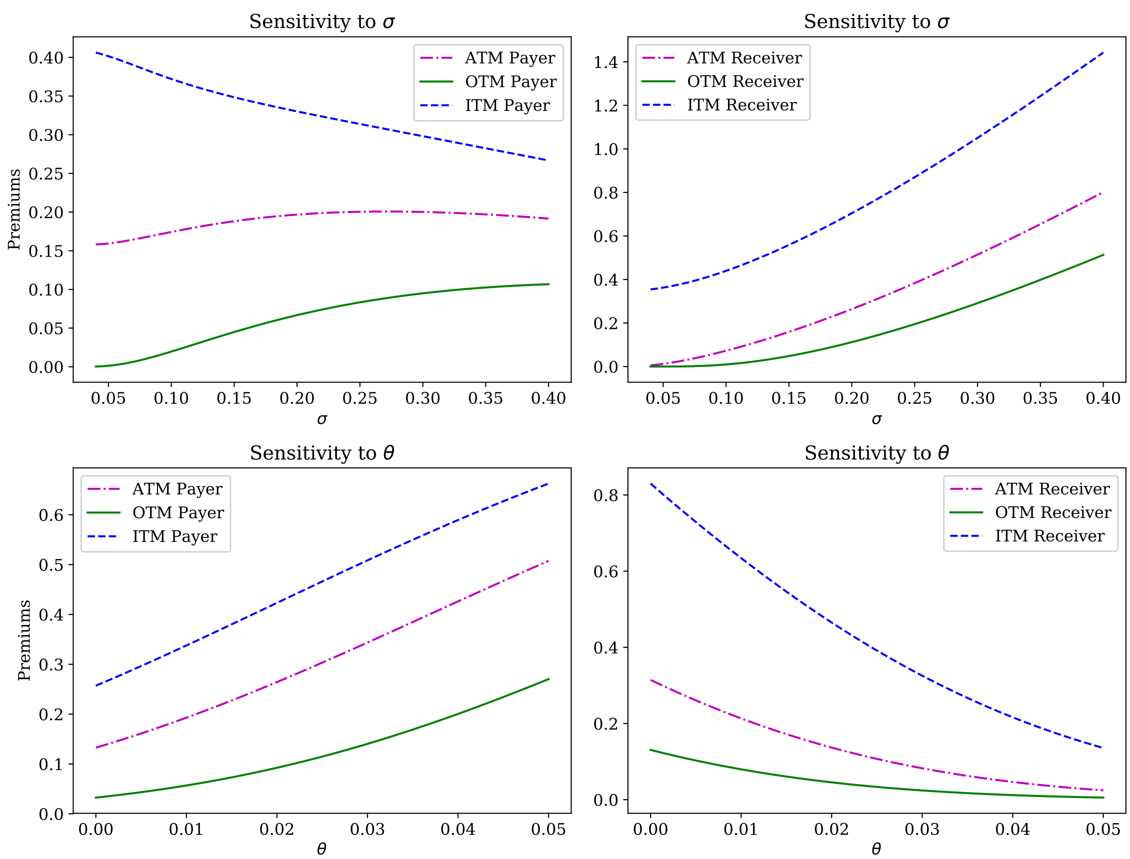
$$K^* = \frac{\mathbb{E}^T[A_c(S(T)) \cdot S(T)]}{\mathbb{E}^T[A_c(S(T))]}. \tag{4.11}$$

Another manifestation of the zero-wide collars mispricing issue under the approximation formula is in the pricing of individual ATM payer or receiver. Note that ATM straddles are liquidly quoted in the market, and the approximation formula will lead us to falsely conclude that ATM payer or receiver are both worth exactly half the price of the ATM straddle. The implication of this error can be understood by referring to Equation (2.19), the CMS replication formula. From the equation, It is clear that in order to price CMS products using the static replication approach, it is vital that one knows the price of the ATM payer or receiver swaption, as it is the first instrument that comes into the replication portfolio for a CMS cap and floor, respectively. Using the standard market assumption (2.15) will result in incorrect valuation of the first instrument in the replication portfolio, despite having calibrated the volatility surface to match market quotes accurately, thereby affecting the accuracy of CMS pricing.

4.4 Sensitivities

Here we move on to demonstrate the existence of negative vega under our unified market model. Exhibit VII plots the sensitivities of IRR-settled swaption prices to the unified model parameters σ (top figures) and θ (bottom figures). The top figure clearly shows that it is possible for the vega of IRR-settled payer swaptions to become negative. This is attributed to the fact that when volatility increases, the concave payoff profile due to the IRR-annuity discounting can cause the benefit of a larger volatility to diminish for IRR-settled payer swaptions which are already sufficiently in-the-money. Naturally, this characteristic is not observed for IRR-settled receiver swaptions, as increments in volatility coupled with the convex payoff profile due to the IRR-annuity can only be favourable, given that it increases the value of the swaption. The existence of negative vega, i.e. negative volatility sensitivity, has been reported earlier in Tee & Kerkhof (2014) and Bermin & Williams (2017).

Exhibit VII: Sensitivities of model parameters.



As a summary, we tabulate the differences between the swap market model and our unified model across different product types in Exhibit VIII. The * symbol in the table under the swap market model column denotes the use of the approximation formula, which leads to pricing discrepancies.

Exhibit VIII: Comparison of swap market model *vs* unified market model.

	Swap Market Model	Unified Market Model
Calibration	swaptions only	swaptions (both settlement types) + CMS
Market swap-settled swaptions	closed-form	closed-form
Market IRR-settled swaptions	closed-form*	1-d integral
CMS products	1-d integral	closed-form
Zero-wide collars	0*	1-d integral
ITM IRR-settled swaptions	put-call parity*	1-d integral

5 Conclusions

In this paper, we have derived a general model to price both IRR-settled and swap-settled swaptions, CMS products, and zero-wide collars in a consistent manner, This has been achieved by formulating a unified framework which price swaptions of either settlement types and CMS products under a single market model. The model is able to capture the convex and concave payoff profiles of IRR-settled swaptions, and evaluate the expectations under a consistent risk-neutral measure.

Practitioners have a strong preference to use a market model to price liquid products. Although a market model based on Black formula is able to handle swap-settled swaptions consistently, two simplifying assumptions are required to extend it to handle IRR-settled swaptions. This gives rise to the zero-wide collar pricing problem, the ITM IRR-settled swaptions pricing problem, and the negative vega problem. Our unified model is able to resolve all three problems while maintaining a sufficient degree of analytical tractability.

We have shown that contrary to swap-settled swaptions, when the strike is equal to the forward swap rate, the values of IRR-settled payer and receiver swaptions are different, which the approximation formula used widely in the industry fail to capture. In addition, the implied volatilities of IRR-settled payers and receivers are different when we do not make the two simplifying assumptions used to derive the approximation formula. We have also demonstrated that using the approximation formula leads to mispricings of ITM IRR-settled swaptions. These findings offer important insights to the ongoing reform in the European swaption market.

Notes

[1]European swaption market convention has changed to use “colateralized cash price”, i.e. swap-settled with OIS discounting. Dealers and brokers have committed to quote both IRR-settled and swap-settled swaption prices. See `https://www.isda.org/2018/11/26/market-practice-change-for-settlement-of-eur-swaptions-to-collateralized-cash-price/`

References

Andersen, L. B. G. & Piterbarg, V. (2010), *Interest Rate Modeling. Volume III: Products and Risk Management.* Atlantic Financial Press.

Bakshi, G., Kapadia, N. & Madan, D. (2003), 'Stock return characteristics, skew laws and the differential pricing of individual equity options', *Review of Financial Studies* **16** (1), pp. 101–143.

Bermin, H. P. & Williams, G. (2017), 'On cash settled IRR-swaptions and Markov functional modeling', *International Journal of Theoretical and Applied Finance* **20** (2), pp. 1–20.

Bianchetti, M. & Carlicchi, M. (2013), 'Markets evolution after the credit crunch', *Interest rate modelling after the financial crisis*, ed. by M. Bianchetti & M. Morini, London: Risk Books.

Black, F. (1976), 'The pricing of commodity contracts', *Journal of Financial Economics* **3** (1-2), pp. 167–179.

Boenkost, W. & Schmidt, W. M. (2009), 'Interest Rate Convexity and the Volatility Smile', *Working Paper*, [Online] Available at: `http://ssrn.com/abstract=1375581`.

Brace, A., Gatarek, D. & Musiela, M. (1997), 'The market model of interest rate dynamics', *Mathematical Finance* **7** (2), pp. 127–155.

Breeden, D. T. & Litzenberger, R. H. (1978), 'Prices of state-contingent claims implicit in option prices', *Journal of Business* **51**, pp. 621–651.

Brigo, D. & Mercurio, F. (2006), *Interest rate models – theory and practice: with smile, inflation and credit*, New York: Springer.

Carr, P. & Madan, D. (2001), 'Towards a theory of volatility trading', *Handbooks in Mathematical Finance: Option Pricing, Interest Rates and Risk Management*, ed. by E. Jouini, J. Cvitanic & M. Musiela, Cambridge University Press, pp. 458–476.

Cedervall, S. & Piterbarg, V. (2012), 'Full implications for CMS convexity', *Risk* (April), pp. 45–51.

De Jong, F., Driessen, J. & Pelsser, A. (2001), 'Libor market models versus swap market models for pricing interest rate derivatives: an empirical analysis', *European Finance Review* **5** (3), pp. 201–237.

Feldman, K. (Mar. 2020), 'Change of measure in midcurve pricing', *Wilmott* (106), pp. 76–81.

Feldman, K. & Portheault, N. (2017), 'How much is your zero collar worth?', *Automated Trader* **42** (1), pp. 45–49.

Hagan, P. (Mar. 2003), 'Convexity conundrums: pricing CMS swaps, caps and floors', *Wilmott Magazine*, pp. 38–44.

Hagan, P., Kumar, D., Lesniewski, A. & Woodward, D. (Sept. 2002), 'Managing Smile Risk', *Wilmott*, pp. 84–108.

Heath, D., Jarrow, R. & Morton, A. (1992), 'Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation', *Econometrica* **60** (1), pp. 77–105.

Henrard, M. (2010), 'Cash-settled swaptions: how wrong are we?', *OpenGamma Working Paper*, [Online] Available at: `http://ssrn.com/abstract=1703846`.

Hull, J. & White, A. (1987), 'The pricing of options on assets with stochastic volatilities', *Journal of Finance* **42** (2), pp. 281–300.

Hull, J. & White, A. (1990), 'Pricing interest-rate derivative securities', *Review of Financial Studies* **3** (4), pp. 573–592.

Hunt, P. J. & Kennedy, J. E. (2000), *Financial Derivatives in Theory and Practice*, England: John Wiley & Sons.

Jamshidian, F. (1997), 'LIBOR and swap market models and measures', *Finance and Stochastics* **1** (4), pp. 293–330.

Jiang, G. & Tian, S. (2005), 'The model-free implied volatility and its information content', *Review of Financial Studies* **18** (4), pp. 1305–1342.

Lee, R. & Wang, D. (2012), 'Displaced lognormal volatility skews: analysis and applications to stochastic volatility simulations', *Annals of Finance* **8** (2-3), pp. 159–181.

Lutz, M. (2015), 'Two collars and a free lunch', *Working Paper*, [Online] Available at: https://ssrn.com/abstract=2686622.

Mercurio, F. (2008), 'Cash-settled swaptions and no-arbitrage', *Risk* (February), pp. 96–98.

Miltersen, K. R., Sandmann, K. & Sondermann, D. (1997), 'Closed form solutions for term structure derivatives with log-normal interest rates', *Journal of Finance* **52** (1), pp. 409–430.

Pietersz, R. & Sengers, F. (2017), 'Cash-settles swaptions: a new pricing model', *Working Paper*, [Online] Available at: https://ssrn.com/abstract=3041228.

Pugachevsky, D. (2001), 'Foward CMS rate adjustment', *Risk* (March), pp. 125–128.

Rubinstein, M. (1983), 'Displaced diffusion option pricing', *Journal of Finance* **38**, pp. 213–217.

Svoboda-Greenwood, S. (2009), 'Displaced diffusion as an approximation of the constant elasticity of variance', *Applied Mathematical Finance* **16** (3), pp. 269–286.

Tee, C. W. & Kerkhof, J. (2014), 'A payoff consistent approach to cash-settled swaptions and CMS replication', *Working Paper:* 8^{th} *World Congress of Bachelier Finance Society, Brussels.*

Zhang, M. & Fabozzi, F. J. (2016), 'On the estimation of the SABR model's beta parameter: the role of hedging in determining the beta parameter', *Journal of Derivatives* **24** (1), pp. 48–57.

Zimmermann, H. (2018), 'The second partial derivative of option price with respect to the strike: a historical reminiscence', *Journal of Derivatives* **25** (3), pp. 81–87.