Department of Economics, Finance & Accounting

Working Paper N305-20

Identifying Statistical Arbitrage in Interest Rate Markets: A Genetic Algorithm Approach

J. C. Arismendi-Zambrano,[*] T. Ramos-Almeida,[†] J. C. Reboredo,[‡] and M. A. Rivera-Castro[§]

June 2020

Abstract

In this paper a multidimensional term structure model is used to find statistical arbitrage opportunities in the interest rates derivatives market. The implied volatility of the model is calibrated by using a genetic algorithm optimization method. Two different options over the same underlying interest rate asset are tested, using data from a weak efficient economy market – Brazilian derivatives market. The results show that there is no systematic mispricing between these two options, but temporary arbitrage opportunities perceptible to the average informed trader are possible.

JEL classification: G13, C51, C52, C63

Keywords: Interest Rates Derivatives, Financial Economics, Arbitrage, Swaption-Cap Puzzle

[*]Department of Economics, Finance and Accounting, National University of Ireland, Maynooth, Ireland. ICMA Centre, Henley Business School, University of Reading, Reading, UK. email: juancarlos.arismendizambrano@mu.ie, j.arismendi@icmacentre.ac.uk

[†]Federal University of Bahia, UFBA, Salvador, Brazil. e-mail: thiagora2007@gmail.com

[‡]Department of Economics, University of Santiago de Compostela, Santiago de Compostela, Spain. email: juancarlos.reboredo@usc.es

[§]University of Salvador (UNIFACS), Salvador, Brazil. Department of Applied Social Sciences, State University of Feira de Santana, Feira de Santana, Brazil. email: marcastro@uefs.br

I. Introduction

Derivatives are used with three purposes: hedging, speculation and arbitrage. Arbitrageurs provide liquidity and equilibrium to the markets by searching for distortions in financial assets prices with the objective of profiting from any difference between two equivalent assets with different prices; according to Ross (2005) arbitrage occurs when there is an infinite return: a positive return with no investment. Following Backus et al. (1998) we can infer that arbitrage in interest rate options markets is possible due to fundamentals deviations. The *Efficient Market Hypothesis* (EMH) developed by Fama (1970) is fundamental in many results of financial economics theory: in its strong form, no past, present or inside information might affect future prices of assets. However, not all financial markets have the same mechanisms, technology, and speed of information transmission; for example, emerging markets. Lo and MacKinlay (1988), Lo and Wang (1995), and Lo (2004), proposed the *Adaptive Market Hypothesis* (AMH), where some assumptions of the EMH are relaxed. Urrutia (1995) found that Latin America equity markets are not efficient, but weak-form efficient. Al-Khazali and Mirzaei (2017) find that AMH is supported in Islamic stock indices, but they gained efficiency over time. Ansotegui et al. (2013) identified statistical arbitrage in a pairs-trading strategy in the Argentinean and Egyptian equity markets, by exploring the depository receipts vs. the stock parity.[1]

Among the emerging economies derivatives markets, B3, the Brazil Stock Exchange and Over-the-Counter Market, represents the second largest by number of transactions – 3.88 billion, measured by the Futures International Association,[2] and the third largest in the world, with an equivalent to 80% of the total transactions of the Chicago Mercantile Exchange (CME) group.[3] One of the most traded derivative is the one-day (overnight) inter-bank deposit rate futures contract (first instrument of interest). The one-day inter-bank deposit futures contract dynamic is similar to an interest rate swap contract, one counter-party assuming the floating rate risk, and the other counter-party assuming the fixed rate risk.

In this paper we used an arbitrage-free model to test mispricing in the B3 Brazilian interest rate derivative market: we test two derivatives which have as the underlying variable the one-day (overnight)

[1] In the case of developed markets, intraday parity relationships can also provide arbitrage opportunities; for instance, Alsayed and McGroarty (2012) and Mitra et al. (2019) find arbitrage possibilities between the American depository receipts (ADRs) and their corresponding stocks, for a group of UK companies.

[2] https://www.fia.org/resources/global-futures-and-options-trading-reaches-record-level-2019.

[3] Brazil is the third largest emerging economy by GDP, and the eighth largest in the world by a report of the United Nations (2019).

inter-bank deposit rate, henceforth DI interest rate. As expectations of the DI interest rate, needed to test arbitrage, are better approximated by the one-day inter-bank deposit rate future contract, henceforth DI future contract (first instrument of interest), we use it.

We use the one-day interbank deposit future *options* (second instrument of interest), henceforth, *DI futures options*, and the one-day average interbank deposit rate index *options*, henceforth *IDI index options* (third instrument of interest). We find that there exists temporary opportunities for arbitrage in the Brazilian interest rate derivative markets by considering these two options, but these opportunities are not sustainable in the long term. In our approach, we use a genetic algorithm to solve the non-convex stochastic optimization problem that is derived from the non-arbitrage conditions between the two interest rates options. To our knowledge, this is the first time an evolutionary algorithm has been used to solve an interest rates option arbitrage mispricing problem.

The contributions of this paper to the literature are twofold: First, we provide an empirical analysis of the arbitrage conditions of the two option instruments considered:[4] the existence of arbitrage might weaken the EMH hypothesis of Fama (1970), and support the AMH of Lo (2004), having implications for the asset pricing theory: by EMH there should be no arbitrage opportunities, or only a small set of opportunities imperceptible to the average informed trader in our arbitrage framework. This empirical analysis of the behavior of the interest rates option prices and its arbitrage opportunities is an important channel to identify how rational the arbitrageurs are – or, at least a channel to identify behavioral patterns from market imperfections.[5] The data used for the study is composed of 61 weeks of interest rate options prices from November 03, 2017 to January 18, 2019, and 8.396 different option instruments (by including all different maturities). To estimate the option prices we model the overnight interest rates (DI) with the Longstaff et al. (2001a) model (henceforth the LSS model), developed to simulate the evolution of the term structure of the London Interbank Offered Rate (LIBOR) in continuous time. The genetic algorithm

[4]By option pricing theory results (Merton et al., 1973), the relationship between these two options are driven primarily by the correlation structure of the forward rates, that forces the no-arbitrage relationship. The first option pricing interest rate models that were arbitrage-free (Ho and Lee, 1986; Ritchken and Boenawan, 1990) considered a mathematical implicit no-arbitrage relationship by exploiting the implicit expectation rates in two discrete periods; nevertheless, they were consistent only with the spot interest rate.In some applications with discrete time, one-day overnight interest rate can be considered a spot rate. Nevertheless, in our paper we consider a more general continuous time model. Heath et al. (1992) extended Ho and Lee's (1986) model to the entire term structure, and it became a reference in interest rate option pricing; but Amin and Morton (1994) and Backus et al. (1998) found that arbitrage-free models might misprice some interest rate options when the fundamental and stylized factors – such as mean reversion – are not considered.

[5]In line with Xia (2001), for interest rates options traders, the learning curve of the arbitrage opportunities existing in the market might be long enough to make the arbitrage opportunities perceptible under our framework.

method is used to capture the average implied volatility of the options. The genetic algorithm is selected given that the optimization function is non-convex. The algorithm's objective is to find the Black (1976) implied volatility that best prices a set of options.

Previously in the literature, evolutionary algorithms have been used to solve statistical arbitrage in equity markets (Burgess, 1999, 2005). Other artificial intelligence techniques were used to solve statistical arbitrage, pairs-trading, and portfolio/asset management problems: Hung et al. (1996) used neural networks to solve statistical arbitrage factors by applying the arbitrage pricing theory (APT) in equities, and more recently, Krauss et al. (2017) used deep neural networks to extract statistical arbitrage opportunities in the S&P stock index. Krauss (2016) presents a survey of the different meta-heuristics approaches for solving the statistical arbitrage problem. Figlewski (2016) provides a deep analysis of the conditions under which arbitrage is possible; between the four arbitrage examples he analyzed, our study fits into the second category, where "arbitrage is easy only in theory", given that there might exists limitations in the liquidity of the options. Nevertheless, our contribution to the theory is to develop a methodology to identify where that *possible arbitrage* might occur.

Our empirical results provide evidence that there is no systematic arbitrage opportunities[6] between the *IDI index options* and the *DI futures options*. Nonetheless, for some periods of the sample, there were some significant mispricings, that can lead to temporary profits with statistical arbitrage. Our empirical analysis contributes to the literature on interest rates markets efficiency: Dunis and Lequeux (2000) analyzed the spread between the bond prices and the implicit hedge price of the corresponding futures in the European bond markets, finding that intra-day data contributes towards the reduction of the hedge ratio variance. Choi et al. (2017) studied the variance risk premium of US treasury bonds, and find that the variance risk premium is persistent and economically significant, which will represent a potential source of statistical arbitrage. We contribute also to the general theory of statistical arbitrage, as our methodology can be extended to other assets with interest rates. For example, complex models of equity statistical arbitrage incorporate the dynamics of the interest rates. Statistical arbitrage in equity markets is reviewed and

[6]Recent models, such as Filipović and Trolle (2013) considered the arbitrage opportunities in previous arbitrage-free models, and derived an option pricing model that disentangled the misprices into liquidity (non-default) and default components. Filipović and Trolle (2013) model is relevant to our work as they considered a reduced-form model over the swaps on the LIBOR and the overnight interest rate, that could be used to disentangle the source of the temporary statistical arbitrage, if it is due to default risk or due to liquidity risk, as first explained by Liu et al. (2006). The Filipović et al. (2017) model is also relevant to our work, as some extensions related to our results might be derived from the linear rational condition used in their model.

developed in Figlewski (1989); Bondarenko (2003); Alexander and Dimitriu (2005), and Avellaneda and Lee (2010).

Our second contribution is methodological: by using a genetic algorithm optimization apporoach, we provide an alternative test for statistical arbitrage in interest rate derivatives markets. The genetic algorithm finds the implied volatility of the Black (1976) option pricing model that adjusts to the observed empirical data, where the interest rate term structure for valuation follows the Longstaff et al. (2001a) (LLS) string model. To gain computational tractability and simplify the equations, we changed the discrete time interest rate capitalization equations into continuous time capitalization equations. A significant part of the literature about interest rates derivatives, in accordance with Jarrow et al. (2007), considers two subjects: (i) the first one is related to the unspanned stochastic volatility, the risk factors that pricing interest rates derivatives are not the same as the term structure, and (ii) The second one, is about the relative valuation of Caps e Swaptions.[7] Given that the two options under study have the same underlying asset, prices movements of both options are correlated to the movements of the term structure of the DI interest rate, this involves a no-arbitrage relation between these two derivatives. However, according to Longstaff et al. (2001a) and Jagannathan et al. (2003), there is an important mispricing between caps and swaptions – verified through the use of various multifactor term structure models – that has been called the "swaptions/caps puzzle".[8] Our paper contributes to the literature by providing a framework that tests statistical arbitrage between the two options: *IDI index options* and *DI future options*; as such, it contributes to the "swaptions/caps" puzzle, having the IDI index option similarities with a *cap* and the *DI future options* similarities with a *swaption*.

The rest of this work is organized as follows. Section 1 presents the financial instruments utilized in the empirical investigation and its payoffs equations. Section 2 develops the string market model adapted for the DI interest rates. Section 3 reports the dataset used, Section 4 the empirical findings, and we conclude in Section 5.

[7]The swaption is an option that confers on his owner the right to enter in a swap contract of interest rates, and is divided into two types: the *payer swaption* and the *receiver swaption*. With the *payer swaption*, the buyer has the right to enter in a swap contract to receive the floating rate and pay the fixed rate. The *receiver swaption* is the opposite. The *Caps* is a derivative for which the buyer receives a payment at the end of each period in which the interest rate exceed the strike of the option agreed at the moment of the purchase.

[8]Jarrow et al. (2007) developed a model with stochastic volatility and jump-diffusions over the LIBOR forward rates, where the *cap* and *swaption* prices were arbitrage-free.

II. Interbank Deposit Rates (DI) Derivatives Market

A. One-day Interbank Deposit Futures – First instrument

One-day interbank deposit futures, known as *DI futures*, is a derivative contract that has as the underlying the expected one-day interbank loan interest rate, averaged between the trading date and the maturity date of the contract. The product has a notional value of R$ 100,000.00 at maturity, each contract corresponds to 100,000 points and each point amounts to R$ 1.00 at maturity. The futures price reflects the expectations of the future behavior of the one-day interbank deposit rate[9] verified during the period between the initial trading date and the last trading date of the contract. The DI futures contract is usually used to hedge and manage risk exposures on liabilities and debt positions in Brazilian reais by companies. The DI rate is based on interbank loan transactions, and it is the equivalent Brazilian LIBOR rate. At one end, DI futures are very liquid and represent a reference for one-day interbank loans. At the other end, DI futures is one of the most volatile contracts, which will bring many opportunities for speculation and arbitrage opportunities across the entire term structure of interest rates.

One of the main characteristics of the DI futures contract is the margin account, that is used to cover the eventual losses of the operation. To buy a future contract of any commodity or index, there is no a cost at the moment of the purchase, therefore, it is necessary to deposit a guaranteed amount that will be used to control the margin and settlement risk. Define the initial date of the trade t, and the maturity of the future contract T, the daily cash flow of the margin account can be expressed as,

$$MCF_t^T = CP_t^T - TP_t^T, \tag{1}$$

where MCF_t^T is the *margin account cash-flow* for the future contract accounted at date t with maturity in T, but paid on the next day (margin accounts are adjusted daily), CP is the *contract unitary price* (UP) at the date of trading entry t in the future contract, adjusted by the spot rate $r(t)$, and TP is the *contract unitary price* (UP) at the date of entry t in the future contract, but adjusted by the DI rate between date t and the date T.

Consider a continuous trading economy with a continuum of default-free bonds, traded with different

[9]DI rates calculated by CETIP – Custody and Settlement – and expressed as a percentage rate per anuum compounded daily based on a 252-day year.

maturities. Assume bonds are traded at dates $t \in [0,T]$. Define $P(t,T)$ as the bond's price at time $t \in [0,T]$, that has a face value of R\$ 1,00 at maturity T. Henceforth, $P(T,T) = 1$ and $P(t,T) > 0$ for all $t \in [0,T]$ (no default-free condition of the bonds).

The unit price (UP) of the one-day interbank deposit futures contract is defined by,

$$UP(t,T) = \frac{100,000}{(1 + k/100)^{(T-t)/252}}, \tag{2}$$

where k is the agreed interest rate; the difference $T - t$, represents the reserve days,[10] comprised between the trading date t, inclusive, and the maturity date of the contract T, exclusive. DI futures contracts have a value of R\$ 100,000.00 at maturity, then, the UP represents the R\$ 100,000.00 discounted by the agreed interest rate.[11] The bank calendar in Brazil uses annual periods of 252 business day and the capitalization of the DI futures contract occurs only at the business day verified between the trade day and the maturity date, inclusive.

Define two maturity dates, T_i and T_j, the *forward interest rate* $f(t, T_i, T_j)$ is the implicit interest rate between T_i and T_j, but discounted in t, and its price is calculated as

$$f(t, T_i, T_j)^{\text{annual}} = \left(\frac{UP(t, T_i)}{UP(t, T_j)} \right)^{\frac{252}{T_j - T_i}} - 1, \quad \forall T_i, T_j \in [t, T], T_i \leq T_j, t \in [0, T], \tag{3}$$

where $UP(t, T_i), UP(t, T_j)$ are the unit prices at time t of the contracts that have maturity at T_i and T_j correspondingly.

To simplify the equations and facilitate the numerical implementations of the model, we use all interest rates with continuous time capitalization. Therefore, we use the following expressions to get the annual forward rate with continuous time capitalization

$$f(t, T_i, T_j)^{\text{continuous}} = \log\left(1 + f(t, T_i, T_j)^{\text{annual}}\right), \quad \forall T_i, T_j \in [t, T], T_i \leq T_j, t \in [0, T]. \tag{4}$$

Because the calendar of DI future maturities is discrete and discontinuous – only 12 maturities in a

[10] Reserve days are business days for the purposes of financial market transactions, as established by the National Monetary Council.

[11] We need to provide the full details of the DI future contract, as by being a one-day contract, 1/252 fraction of a year represents a full period of trading.

252-business days year, that represents 4.76% of the full trading calendar – we apply an interpolation technique commonly used in the literature and by practitioners: we perform a piecewise polynomial interpolation (cubic spline) to calculate a smooth curve of the interest rates to fill the gaps for the remaining 95.24% of the days.

With the DI rates for the diverse maturities of the DI futures contract, $f(t, T_i, T_j)$, one can obtain the bond prices that pay R$ 1,00 at the maturity,

$$P(t, T_j) = \frac{1}{\exp\left(f(t, t, T_j)\frac{T-t}{252}\right)}, \quad \forall T_j \in [t, T], t \in [0, T]. \tag{5}$$

where $r(t)$ is the spot rate,

$$r(t) = f(t, t, t), \quad \forall t \in [0, T]. \tag{6}$$

The DI futures contract will have two legs, a fixed (PV_{fixed}), and a floating (PV_{float}), that should be equal, $PV_{\text{fixed}} = PV_{\text{float}}$. By definition,

$$\begin{aligned} PV_{\text{fixed}} &= \sum_t C\delta_t PO_t^{\text{fixed}}\Lambda_t = C\sum_t \delta_t PO_t^{\text{fixed}}\Lambda_t, \\ PV_{\text{float}} &= \sum_t r_t\delta_t PO_t^{\text{float}}\Lambda_t, \end{aligned}$$

with PO_t^{fixed} the net present value at time t of the overnight interest rate agreed in the future contract, PO_t^{float} the net present value of the expected effective DI interest rate (that will be an average of the traded rate until the maturity of the contract); C, the contracts interest rate (fixed leg); Λ_t - the discount factor for payment date t and r_t - the spot rate (floating rate of future payment); and δ_t - the day count fraction. Then,

$$C = \frac{\sum_t r_t\delta_t PO_t^{\text{float}}\Lambda_t}{\sum_t \delta_t PO_t^{\text{fixed}}\Lambda_t}, \tag{7}$$

B. One-Day Interbank Deposit Futures Options - Second instrument

At the exercise day of the futures contract (maturity day), the buyer and the seller receive a position in the one-day interbank deposit futures contract. The buyer receives a long position and the seller receives a short position on the futures contract.

Let T_1 be the exercise date of the option and T_2 the maturity date of the future contract; the option's *payoff* can be expressed in T_1 by the following manner,

$$\text{payoff}\left[T_1\right] = Q \times \max\left(cp\left[\frac{100,000}{\exp\left(k\left(\frac{T_2-T_1}{252}\right)\right)} - FUT_{T_1,T_2} \right], 0 \right), \tag{8}$$

where Q is the contract quantity; cp is the variable that defines if the option is of the type *call* or *put*; and FUT_{T_1,T_2} is the notional value of the future contract at time T_1. The future contract will have a notional value of R$ 100,000 at maturity (T_2).

Under the classification of the B3 stocks and derivatives exchange, the maturity of the option depends on the type of the future: (i) **Type 1 (D11)**: when the underlying asset of the option is the future contract with maturity in 3 months after the maturity of the option; (ii) **Type 2 (D12)**: when the underlying asset of the option is the future contract with maturity in 6 months after the maturity of the option; (iii) **Type 3 (D13)**: when the underlying asset of the option is the future contract with maturity in 1 year after the maturity of the option; (iv) **Type 4 (D14)**: when the underlying asset of the option is the future contract with maturity specified by the exchange securities. Given that the underlying asset has an expiration date that is different from the expiration date of the derivative, there exists a basis risk between the two expiration dates, that exposes the investor to the risks of interest rate term structure parallel shifts and slope shifts. It is worth noting that these options are a *European type*, the exercise occurs only at the maturity date, when the price of the DI future contract is greater than the exercise price for call options; or below to the exercise price for put options.

After some algebraic manipulations, it is possible to transform the *payoff* in (8) to the payoff of a *swaption*, in order to reduce the computational estimation burden,

$$\text{payoff}\left[T_1\right] = Q \times \frac{100,000}{\exp\left(k\left(\frac{T_2-T_1}{252}\right)\right)} \times \max\left(cp\left[\frac{\exp\left(r\left(\frac{T_2-T_1}{252}\right)\right) - \exp\left(k\left(\frac{T_2-T_1}{252}\right)\right)}{\exp\left(r\left(\frac{T_2-T_1}{252}\right)\right)} \right], 0 \right), \tag{9}$$

and r is the 1-year implicit yield (annual interest rate with continuous time capitalization) between T_1 and T_2, and k is the exercise price of the option (published in rates by the B3).

We transform the continuous rates (exponential) into arithmetic,

$$1 + K^{\text{linear}}\left(\frac{T_2 - T_1}{252}\right) \quad = \quad \exp\left(k\left(\frac{T_2 - T_1}{252}\right)\right), \tag{10}$$

$$1 + R^{\text{linear}} \left(\frac{T_2 - T_1}{252} \right) = \exp \left(r \left(\frac{T_2 - T_1}{252} \right) \right). \tag{11}$$

Substituting (10) and (11) in (9) yields,

$$\text{payoff}\,[T_1] = Q \times \frac{100,000}{\exp\left(k\left(\frac{T_2 - T_1}{252}\right)\right)} \times \left(\frac{T_2 - T_1}{252} \right) \times \max\left(cp \left[\frac{R^{\text{linear}} - K^{\text{linear}}}{\exp\left(r\left(\frac{T_2 - T_1}{252}\right)\right)} \right], 0 \right). \tag{12}$$

A final step consists in obtaining the *payoff* at T_2,

$$\text{payoff}\,[T_2] = Q \times \frac{100,000}{\exp\left(k\left(\frac{T_2 - T_1}{252}\right)\right)} \times \left(\frac{T_2 - T_1}{252} \right) \times \max\left(cp \left[R^{\text{linear}} - K^{\text{linear}} \right], 0 \right). \tag{13}$$

Interest rate R^{linear} represents the implicit linear DI rate between T_1 and T_2, K^{linear}, is the implicit linear *strike* rate of the option.

C. One-Day Average Interbank Deposit Rate Index (IDI) Options - Third Instrument

The third instrument (equivalent to the *cap* option) has as underlying the one-day average interbank deposit rate index (IDI). The IDI index is an average of the one-Day interbank deposit rate (DI rate), with daily updates.[12] The index price is calculated as,

$$IDI_T = IDI_{\text{basedate}} \prod_{t_i = \text{basedate}}^{T} (1 + CDI_{t_i})^{\frac{1}{252}}. \tag{14}$$

This index in (14) resembles the caplet and floors instruments, used to protect the buyers against fluctuations in interest rates. Unlike the future contract, there is no cash-flow exchange between the counterparts.

The payoff of an IDI option with maturity at date T, is given by,

$$\text{payoff}_{T+1*} = Q \times \max\left(cp\left[IDI_T - K \right], 0 \right), \tag{15}$$

where Q is the quantity in number of contracts; K the option strike, and cp is the variable that defines if the option is a *call* or a *put*. It is worth highlighting that the DI rate of any day is released at the end of

[12] The base date of the index is February, 01, 2009, and the base value at this date was R$ 100,000.00.

the same day, therefore, the index price at T can be known only in $T+1$.

The payoff also can be expressed with a continuous time interest rate,

$$\text{payoff}_{T+1*} = Q \times \max\left(cp\left[IDI_t \exp\left(r\left(\frac{T-t}{252}\right)\right) - K\right], 0\right). \tag{16}$$

III. Methodology

A. *Valuations Models*

The classical model for interest rates options valuation used by the industry is the Black (1976) model, where the price is quoted in terms of the implied volatility. In our framework, assuming linear interest rates between T_1 and T_2, and that R^{linear} has log-normal distribution, the *DI futures options* (the second instrument of reference) Black (1976) model price is,

$$C_t = D_{t,T_2}^{CDI} \frac{100,000}{\exp\left(k\left(\frac{T_2-T_1}{252}\right)\right)} \frac{(T_2-T_1)}{252}\left(R_{T_1,T_2}^{\text{linear}} N(d1) - K_{T_1,T_2}^{\text{linear}} N(d2)\right), \tag{17}$$

And,

$$P_t = D_{t,T_2}^{CDI} \frac{100,000}{\exp\left(k\left(\frac{T_2-T_1}{252}\right)\right)} \frac{(T_2-T_1)}{252}\left(K_{T_1,T_2}^{\text{linear}} N(-d2) - R_{T_1,T_2}^{\text{linear}} N(-d1)\right), \tag{18}$$

with,

$$\begin{aligned} d1 &= \frac{\ln\left(\frac{R_{T_1,T_2}^{\text{linear}}}{K_{T_1,T_2}^{\text{linear}}}\right) + 0.5\sigma^2 (T_1 - t)}{\sigma\sqrt{(T_1-t)}}, \\ d2 &= d1 - \sigma\sqrt{\left(\frac{T_1-t}{252}\right)}, \end{aligned}$$

where,

- C_t and P_t is the call and put option price, respectively.
- D_{t,T_2}^{CDI} is the discount factor between the date t and T_2, based on the curve of the DI future,

- k is the option strike, defined in rate;
- σ is the volatility;
- $T_2 - T_1$ are the contract term in the business day, considering the bank calendar,
- $K^{\text{linear}}_{T_1,T_2} = \dfrac{\left(\exp\left(\left(\log\left(1+k\right)\frac{(T_2-T_1)}{252}\right)\right)-1\right)}{\frac{(T_2-T_1)}{252}}$, and,
- $R^{\text{linear}}_{T_1,T_2}$ are the linear interest rates between T_1 and T_2.

For estimating $R^{\text{linear}}_{T_1,T_2}$, we need to adjust the bond price present value to T_1,

$$P(T_1, T_2) = P(t, T_2)\exp\left(f(t,t,T_1)\left(\frac{T_2 - T_1}{252}\right)\right). \tag{19}$$

Then, the implicit interest rate with continuous capitalization, between T_1 and T_2 is,

$$r^{\text{continuous}}_{T1,T_2} = \log\left(\frac{1}{P(T_1,T_2)}\right)\frac{252}{(T_2 - T_1)}. \tag{20}$$

and the corresponding linear interest rate,

$$R^{\text{linear}}_{T_1,T_2} = \left(\exp\left(r^{\text{continuous}}_{T1,T_2}\left(\frac{T_2-T_1}{252}\right)\right)-1\right)\frac{252}{T_2 - T_1}. \tag{21}$$

B. Implied Volatility – Genetic Algorithm

Using the Black model, is possible to get the implied volatility from the traded prices: by inverting Equation (18) we can obtain the implied volatility σ from the option prices C_t quoted in the market. We applied a genetic algorithm to find the inverse of (18). The genetic algorithms optimization process has three main steps: selection, crossover, and mutation.

C. String Market Model

To model the dynamic of the term structure of DI interest rates, we adapted the Longstaff et al. (2001b) (LSS) multifactor model. When single factor arbitrage-free specifications are applied in a multifactor economically motivated framework, they can generate sub-optimal hedging strategies with arbitrage opportunities (Backus et al., 1998). Single factor models, such as the widely used Black (1976) model, are unable to capture the implied volatility smile.

To capture the volatility smile, Goldstein (2000), Longstaff et al. (2001a) and Santa-Clara and Sornette (2001) modeled the evolution the interest rate term structure as a *stochastic string.* In this approach, which is a generalization of the Heath et al. (1992) model, it is possible to generate a dynamic shape of the interest rate term structure. The innovation of the LSS model consists of each of the forward rates $f(t, T_i, T_j)$ behaving as a dependent random variable, that has its own particular dynamic, but with a unique correlation structure between them.

In the LSS model, the evolution of the forward rate on the risk-neutral measure is,

$$dF_i = \alpha_i F_i dt + \sigma_i F_i dZ_i, \tag{22}$$

where α_i is an unknown *drift*, σ_i is a deterministic volatility function of the forward price, and dZ_i is a Brownian motion process.

Although the model is specified in terms of the forward rates, we consider a more efficient implementation using the vector of the bonds prices. The forward rate with continuous time capitalization are,

$$F_i = \log\left(\frac{P(t, T_i + \tau)}{P(t, T_i)}\right)\frac{\tau}{252}, \tag{23}$$

where τ denotes a certain portion of the business day. Given that we choose a discretization of the data based on the difference between the expiration date of the DI futures, τ express the difference in business days between the maturity $T_i + \tau$ and T_i.

In the LSS model the evolution of the bonds prices is based in a stochastic differential equation with two terms, the *drift* extracted from the spot interest rates, and the *diffusion* composed by the multiplication of: (i) a Jacobian matrix of the sensitivities of the forward rates to each pair correlations, (ii) a matrix of the implied covariance, (iii) a vector of forward rates. Applying Itô's rule to the vector P of the bond prices, we obtain the risk-neutral measure bond price,

$$dP = r(t)Pdt + J^{-1}\Sigma_t F dZ, \tag{24}$$

where,

$$J = \begin{bmatrix} -\frac{P(1)}{P^2(2)} & 0 & 0 & ... & 0 & 0 & 0 \\ \frac{1}{P(3)} & -\frac{P(2)}{P^2(3)} & 0 & ... & 0 & 0 & 0 \\ 0 & \frac{1}{P(4)} & -\frac{P(3)}{P^2(4)} & ... & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{P(29)} & -\frac{P(28)}{P^2(29)} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{P(30)} & -\frac{P(29)}{P^2(30)} \end{bmatrix},$$

$$\Sigma_t = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,N-1} \\ \sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N-1,1} & \sigma_{N-1,2} & \cdots & \sigma_{N-1,N-1}^2 \end{bmatrix},$$

and $r(t)$ is the implicit yield or implicit spot interest rate, $\Sigma_t F dZ$ is a vector defined by the second terms in the Equation (22), and J is the Jacobian matrix obtained through the derivation of the forward vector $(F_1, F_2, \ldots, F_{29})$ with respect to the discount bond price $(P_1, P_2 ... P_{30})$.[13]

The bond prices dynamic in Equation (24) provides a complete specification to the evolution of the interest rate term structure. In Equation (24), $r(t), P, J$, and F can be estimated from the derivatives market prices. The only matrix that is missing in the estimation is Σ_t, the instantaneous variance–covariance matrix, which is estimated by calculating the implied volatility of the traded option prices.

D. Implied Covariance Matrix

The instantaneous variance–covariance matrix Σ_t used in model (24) is calculated in t, with the *DI futures options* traded at the same date t. By definition, a covariance matrix is symmetric and positive semi-definite.[14] We preserve this condition in the estimation procedure. First, we calculate the historical sample correlation H_t, with an increasing rolling window, initialized with 3 years (750 days). Matrix H_t

[13] We use a $(30-1) \times (30-1)$ matrix of sensitivities, as the maximum number of days between two consecutive maturities is less than 30 days.

[14] A matrix is positive semi-definite if, and only if, its eigenvalues are all positive.

can be decomposed as,

$$H_t = U_t \Lambda_0 U_t', \tag{25}$$

where Λ_0 is a diagonal matrix of the eigenvalues of H_t (non-negative), and U_t corresponds to the N eigenvectors. The matrix H_t will represent the estimated correlation in t. To approximate Σ_t from Equation (24), we generate $\hat{\Sigma}_t$, assuming it shares the same eigenvectors of the historic correlations matrix H. Then, $\hat{\Sigma}_t$ can be estimated in the following manner,

$$\hat{\Sigma}_t = U_t \Psi_t U_t', \tag{26}$$

where U_t is the historical correlations eigenvectors observed until t, extracted from Equation (25), and Ψ_t is a diagonal matrix with non-negatives values, that represents the eigenvalues of the factor decomposed historical covariance observed until t. The estimated factor covariance $\hat{\Sigma}_t$ is our approximation for the implicit model covariance Σ_t in Equation (24). The i-th diagonal element of Ψ_t can be interpreted as the instantaneous variance of the i-th factor that drives the interest rates evolution. Considering that the eingevectors of H_t are orthogonal factors, then, the same factors that generate the historic correlation matrix H_t, are used to generate the factor implied covariance matrix, $\hat{\Sigma}_t$.

The diagonal matrix $\hat{\Psi}_t$ estimated values will minimize the root mean squared error (RMSE) of the percent difference between the market prices and model prices of these options (the objective function). Because the RMSE of the market and model price differences are non-convex, we use a genetic algorithm and a cubic interpolation to estimate $\hat{\Psi}_t$.

Using $\hat{\Sigma}_t$, we simulate 2.000 paths of the vector of bond prices by Equation (24). To reduce the variance of the simulated values, we apply antithetic variates.[15] With the matrix of the bond prices simulated, it is possible to find the forward rates between the various maturities using Equation (23). DI future option's prices (second instrument of interest) are yielded by the *payoff* Equation (8).

IDI index option prices (third instrument) are estimated in a similar way, but substituting the payoff (Equation (15)) that is discounted by the average stochastic discount factor (SDF) accumulated among

[15] To maintain the time homogeneity of the model, we excluded the first line and columns of Σ_t and J, this process is repeated up to the maturity of the longer bond.

the dates t and T_2: $\prod_{i=0}^{T_2-1} P(i, i+1)$.

E. The Genetic Algorithm

Let $\Phi(T_j, K^{\text{linear}}, R^{\text{linear}}, \sigma_j, y)$, for $j \in \{1, \ldots, T_N\}$, be non-convex error pricing functions between: (i) the option prices observed in the market (y_i), and (ii) the Black (1976) option prices (Equations (18) and (24)), where $K_i^{\text{linear}} \in \{K_1^{\text{linear}}, \ldots, K_n^{\text{linear}}\}$ are the n available market strikes for call and put options at T_j, and $T_{\max}$ the total number of maturities available in the market; this difference is the standard calibration method used to extract the implicit volatility σ_j. Then, we solve the following stochastic optimization problem,

$$\begin{aligned} \underset{\sigma_j \in \Re}{\text{minimize}} \quad & \hat{F}(\sigma) = E[\Phi(T_j, K_i^{\text{linear}}, R^{\text{linear}}, \sigma_j, y)] \\ &= \frac{1}{n}\left(\left(\sum_{i=1}^{n_1} C_{t,i}(T_j, K_i^{\text{linear}}, R^{\text{linear}}, \sigma_j) - y_i\right)^2 + \left(\sum_{i=n_1+1}^{n} P_{t,i}(T_j, K_i^{\text{linear}}, R^{\text{linear}}, \sigma_j) - y_i\right)^2\right), \end{aligned} \tag{27}$$

where $C_{t,i}(T_j, K^{\text{linear}}, R^{\text{linear}}, \sigma_j) and P_{t,i}(T_j, K^{\text{linear}}, R^{\text{linear}}, \sigma_j)$ are the model price of the *DI futures options* at time t, with maturity in T_j with strike K^{linear}, volatility σ_j, and DI interest rate R^{linear}. N is the number of maturities considered ($N = 30$ in our case). The only unknown values in the minimization problem (27) is the set of implied volatilities (σ_j) that correspond to the volatility in Equation (22); equal to the diagonal of the matrix Σ_t. To find the implied volatility, we apply a genetic algorithm; the first steps consists of generating the initial population set and evaluating this population on the objective function (see Algorithm 1). The second step of the genetic algorithm is to generate a crossover of the population set that has the best performance (see Algorithm 2), and the third step consists in computing the mutation of the population after the crossover (see Algorithm 3).

After the mutation, the population is again evaluated on the objective function. This is an iterative process: the new population is ranked and the best 50% individuals survive the cut, then, the survivors are matched with each other in the crossover to generate descendants. The mutation process will add the "non-convexity" capacity to the search heuristics. The new population is re-evaluated under these three steps until the algorithm reaches a convergence to a global optimum. The mutation step provides an heuristic that can pull the algorithm from a local minimum region, towards a better local minimum region,

Algorithm 1 Evaluating the first population set

Step 1. $\sigma_{j,j}$ = initial volatility
Step 2. Compute the initial population set: $p = (\sigma_1, ..., \sigma_n)$
for k = 2:popsize **do**
  for m=1:nbits **do**
    $\sigma_{m,k} = |\delta|$ where δ is a random number $\delta \sim N(\sigma_{j,j}, 2)$
  end for
end for
Where popsize=20 and nbits=1
Step 3. Evaluate the objective function:
for i=1:popsize **do**
  $cost(i,1) = \frac{1}{n}\left(\left(\sum_{i=1}^{n_1} C_{t,i}(T_j, K_i^{\text{linear}}, R^{\text{linear}}, \sigma_j) - y_i\right)^2 + \left(\sum_{i=n_1+1}^{n} P_{t,i}(T_j, K_i^{\text{linear}}, R^{\text{linear}}, \sigma_j) - y_i\right)^2\right)$
end for

Algorithm 2 Crossover of the population set

Step 1. Sort the population set by the performance on the objective function;
Step 2. Crossover between the best 50% of the population set:
for i = 1:M **do**
  $M = (popsize - keep)/2$ where keep is the number o survivor in the generation
  $dad = \delta * popsize * selection$ where δ is a random number $\delta \sim unif(0,1)$
  $mon = \delta * popsize * selection$ where δ is a random number $\delta \sim unif(0,1)$
  selection=0.5 (fraction of the population kept)
  $indx = 2 * (i-1) + 1$
  $p(\sigma_{1,keep+indx}) = 0.7 * p(\sigma_{1,dad}) + 0.3 * p(\sigma_{1,mon})$
  $p(\sigma_{1,keep+indx+1}) = 0.6 * p(\sigma_{1,dad}) + 0.4 * p(\sigma_{1,mon})$
end for

Algorithm 3 Mutation of the population set

Step 1. Define the quantity of the mutation;
$nmut = popsize * mutrate$
where mutrate is the mutation rate;
for i = 1:nmut **do**
  $col = \delta * popsize \quad \delta \sim unif(0,1)$ is a random number;
  if col is different of 1 **then**
    $pop(1, col) = \delta * p(\sigma_{1,popsize*\delta}) \quad \delta \sim unif(0,1)$ is a random number;
  end if
end for

in the search for the global optimum.

Figure 1 shows the convergence of the genetic algorithm. With less than 15 iterations the algorithm gets to converge to a value of the implied volatility that minimize the objective function.

F. The LSS Model is Arbitrage-Free

In our economy framework, determined by the use of the LSS model, the initial wealth of the investor is R$ 1.00 and that wealth is compounded by the short-term interest rate by,

$$B(t) = \exp\left(\int_0^t r(y)dy\right). \tag{28}$$

Let $Z(t,s) = \frac{P(t,s)}{B(t)}$ be the relative price of the bond.[16] In agreement with Harrison and Kreps (1979), if there exists an equivalent martingale measure P^*, there is no arbitrage opportunities. Within this equivalent risk-neutral probability measure P^*, the relative bond prices $(Z(t,s_1), Z(t,s_2)...Z(t,s_n))$ are martingales.

Consider the system of Equations (29), $b(t,T)$ represents the excess return over the risk-free rate of the bond with maturity T, γ_i represents the risk price of the factor i-th, and $a_i(t,T_j)$ represents the covariance between the bond's return (that has maturity in T_j) with the i-th factor. If there exists an equivalent martingale measure P^*, then, the systems of Equations (29) have a solution; moreover, if the covariance matrix is non-singular, then, the probability measure P^* is unique.[17]

$$\begin{bmatrix} b(t,T_1) \\ \vdots \\ b(t,T_n) \end{bmatrix} = \begin{bmatrix} a_1(t,T_1) & \cdots & a_n(t,T_1) \\ & \vdots & \\ a_1(t,T_n) & \cdots & a_n(t,T_n) \end{bmatrix} \times \begin{bmatrix} \gamma_1(t;T_1,...,T_n) \\ \vdots \\ \gamma_n(t;T_1,...,T_n) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{29}$$

In a world with risk-neutral investors, the risk price that is represented by γ_i is null. Considering that P^* represents the risk-neutral measure, we need to subtract the risk premium from the evolution of the bond prices. In the Equation (24), the *drift* only contains the spot interest rate $r(t)$, considered here as the risk-free rate $r_{RN}(t)$; then, the expected rate of return on all bonds is equal to the spot rate under the

[16] $P(t,s) > B(t)$ because there is a risk premium associated with the tenure of the Bond

[17] For a complete demonstration of this result, see Heath et al. (1992).

risk neutral measure, consequently, the LSS string market model is able to replicate the exact form of the initial term structure and provide an arbitrage-free model.

IV. Data

The data sample of the DI futures contracts consists of weekly observations, from October 01, 2015 to January 31, 2019,[18] and contains 164 weeks, during each week there exists approximately 40 active trading futures, when all the different strikes are considered.

The sample of the *DI futures options* corresponds to the period from November 03, 2017 to January 18, 2019 (61 weeks), with all the options of the sample being in-the-money (ITM) options.[19] The sample contains 2.196 options; from which every week there are 36 options being actively traded. The sample of the *IDI index options* corresponds to the same period of the *DI futures options*, November 03, 2017 to January 18, 2019 (61 weeks). Altogether, 6.100 different options are considered, with approximately 100 options with active trading for each day. All the *IDI index options* are in-the-money (ITM).

Figure 2 shows the surface of the interpolated DI interest rates. Short- and long-term interest rates suffered a shortfall; however, the decline in the long-term rates was significant lower. Central banks usually control the short-term section of the curve with the monetary policy; as a consequence the long-term rates tend to be less susceptible to interventions of the monetary authority.

Figure 3 shows the histogram of the forward rate returns of a particular maturity. These histograms offer an estimate of the underlying empirical probability distribution of the of DI futures contract returns.

Figure 4 shows the interest rate term structure at six different times, between October 01, 2015 and January 31, 2019. In October 23, 2015, the short-term rate was above 13.5%, the curve was upward sloping until the second year, but the long-term rates were reporting a downturn. The macroeconomic scenario was of high inflation – the IPCA (Broad Consumer Prices Index) reached at the end of 2015, 10.67% over the last 12 month. To control inflation, the Central Bank raised the short rate (SELIC) to 14.15% in July 29, 2015. In December 09, 2016, the curve was downward sloping (inverted), that is an uncommon slope situation – sometimes preceding a decrease in the GDP or an economic crisis. The inverted slope denotes

[18] Price observations correspond to the Friday/or last business day of the week after the market has closed.

[19] An option is said to be at-the-money (ATM) if the spot price of the underlying asset $r(t)$ is equal or close to the strike price k, but with a positive payoff.

a market expectation of the interest rates reduction in the medium and the long term. In 2016, the IPCA was of 6.29% and the SELIC closed the year at 13.65%. In 2017 and 2018, the curve again presented the traditional upward slopping form, consistent with a drop in inflation (the IPCA in 2017 was of 2.95% and, in 2018 was of 3.75%) and a drop of the the short-term interest rate – in December 2017, the short-term rate closed at 6.90%, and in 2018, the rate was close to 6.40%.

Figure 5 shows the implicit covariance estimated as in Section III.D. The calendar time and Forward-time axis are equivalent to the term conditions for a *swaption*; for example, 2-year × 5-year point in Figure, represents a *swaption* that gives the right to the owner to enter into an 5-year interest rate swap in two years time. We can observe the 1-year × 1-year and 8-year × 8-year covariance points are higher; a value of 6% means that a change in an hypothetical 8-year × 8-year *swaption* of 1% increases the covariance surface by 6%.

V. Empirical Results

A. *Likelihood Ratio Test*

To estimate the number of factors necessary to price the *DI futures options* with the Longstaff et al. (2001a) string market model, we used an incremental likelihood ratio test. Using the future times series of 61 weeks, with 36 options in each week, we compared nested models with 1 to 5 factors to verify if there were statistical differences among them.[20]

The test is carried in the following manner: for a model with N factors, we compute the sum of squared errors of the difference between the option market and option model weekly prices. Then, the method is applied to a model with $N+1$ factors. The error has a Chi-Square distribution with 61 degrees of freedom (χ^2_{61}), under the one-tail test, and a Chi-Square distribution with 30 degrees of freedom under the two-tail test. hypothesis.

Table I shows the results of the likelihood ratio test between the nested models with 1 to 5 factors. The critical value of χ^2_{61} is 89.59 at the 99% level of confidence, and the critical value of χ^2_{30} is 50.89 for the same level of confidence. The Table I Panel A results show that the relationship is statistically significant

[20]The covariance matrix has N factors when the diagonal matrix Ψ has N eigenvalues across the diagonal.

between the model with one versus two factors, two versus three factors, and three versus four factors. These results imply that the option market prices on DI futures require only fours factors to be priced.

Heath et al. (1990) for the US market, and Almeida et al. (2008) for the Brazilian market, show that three factors (associated with the level, slope and curvature of the interest rates curve) are necessary to make forecasts about the interest rate term structure (IRTS). Therefore, by applying the Longstaff et al. (2001a) model, we find that four risk factors are present on the IRTS. We employ these factors to price the options.

We conduct another likelihood ratio test to get a deeper analysis by dividing the sample into two periods. By virtue of the electoral year with a *ex-ante* uncertain political scenario, the division of the sample is adequate to model the pre-electoral period (first half of the sample), and the post-electoral period (second half of the sample). Even in these two sub-samples there exists evidence about the existence of four factors for modeling the IRTS. Eigenvalues can be interpreted as the implied variance of the factors (Longstaff et al., 2001a), and the eigenvectors as the implied covariances of the factors of the IRTS.

B. String Market Model Performance

Figure 6 shows four sub-plots with the time series for each one of the four eigenvalues of the four factors found in the IRTS, between November 03, 2017 and January 18, 2019. The first eigenvalue, that corresponds to the volatility of the first factor related to the parallel shift of the IRTS, had many statistical arbitrage recognizable patterns over the period analyzed. By the end of 2017, the volatility was very high, and the political scenario was of uncertainty and low confidence due to the lack or delay of structural economy reforms, such as a social security reform.

Curiously, at the beginning of 2018, volatility declined sharply, but after that during the whole of 2018 it started to increase, in line with the electoral campaign for the elections held in October 2018. Even after the election, the volatility continued to increase; analysts deduced that it was due to the uncertain environment concerning the new government's electoral promises.

The volatility of the second eigenvalue, associated with the slope of the interest rate curve, was about 0.2 at the end of 2017, but at the beginning of 2018, suddenly fell to approximately 0.03. In July 2018, closer to the election, the volatility increased to a value similar to the slope's shape at the end of 2017.

Only after the conclusion of the first round of elections is that the volatility felt again.

Third eigenvalue volatility is associated with the curvature of the IRTS (or convexity/concavity). This curvature shape was more pronounced at the end of 2017. In 2018, this value fell significantly and maintained this low level until the end of the sample period. The fourth eigenvalue is related to third order shape properties of the curvature (acceleration of the convexity) and its implied volatility is close to zero. Although it was close to 0.007, during the electoral period this eigenvalue volatility almost doubled to 0.012.

C. Convergence of the Model

Figure 7 shows the root mean squared error (RMSE) of the differences between the option prices obtained from the market and the option prices calculated by the Longstaff et al. (2001a) string market model calibrated to the market prices. We use a four-factor model. The median RMSE on the *DI futures options* sample is 5.8% and the standard deviation is 1.4%. The median RMSE of the *IDI index options* is 3.7% with 1.6% standard deviation. Considering the average *bid–ask spread*[21] and transaction costs that add up to 3% – 4%, we can infer there are no arbitrage opportunities.

D. "Near-The-Money (NTM)" Options and Arbitrage Opportunities

Near-the-money options usually have greater liquidity, therefore traders tend to focus some statistical arbitrage strategies around them. We select a sub-sample with the NTM options and estimate the RSME for both options.

Figure 8 shows the RMSE of the NTM options sub-sample. The RMSE on *IDI index NTM options* is close to 11.35% with 45% standard deviation. The RMSE on *DI futures NTM options* is close to 5.52% with 9.8% standard deviation. We observe that there are cases with possible statistical arbitrage, such as December 2017, or December 2018.

[21]The *bid–ask spread* represents the difference in price that the seller and buyer cast into the trade systems, essentially it is the difference between the highest price at which the buyer is willing to pay and the smallest price at which the seller is willing to sell. The individual that is willing to sell gets the bid and the individual that is willing to buy gets the ask.

VI. Conclusions

In this study we tested the statistical arbitrage opportunities in the derivatives market of a emerging country – Brazilian B3 derivatives market – using a a four-factor Longstaff et al. (2001a) string market model to describe the dynamics of the DI interest rate term structure. We use (i) the *DI futures options* and (ii) the *IDI index options*, which have the same underlying asset – the DI interest rate – to test the statistical arbitrage opportunities. We used a genetic algorithm to calibrate the implied volatility that minimizes the difference between the market and the model prices.

The results show that arbitrage opportunities can exist between the two derivatives. However, there is no arbitrage opportunity of a systematized manner – persistent over time. We found that the "near-the-money" (NTM) options are more likely to present deviations between the model and market prices. Option prices obtained through the Longstaff et al. (2001a) string market model converge to market prices. The average pricing error is close to the typical bid–ask spread values (close to 3-4%), but the "near-the-money" (NTM) sub-sample of options might present some statistical arbitrage opportunities by shorting the *DI futures options* and taking long positions on the *IDI index options*.

Future research can analyze if the deviations that generate arbitrage opportunities are quickly adjusted to fundamental values, or if this deviation is due to the existence of *noise traders* that promote deviations for a the long-run.

References

Al-Khazali, O. and A. Mirzaei (2017). Stock market anomalies, market efficiency and the adaptive market hypothesis: Evidence from islamic stock indices. *Journal of International Financial Markets, Institutions and Money 51*, 190–208.

Alexander, C. and A. Dimitriu (2005). Indexing and statistical arbitrage. *The Journal of Portfolio Management 31*(2), 50–63.

Almeida, C., R. Gomes, A. Leite, and J. Vicente (2008). Movimentos da estrutura a termo e critérios de minimização do erro de previsão em um modelo paramétrico exponencial. *Revista Brasileira de Economia 62*(4), 497–510.

Alsayed, H. and F. McGroarty (2012). Arbitrage and the law of one price in the market for american depository receipts. *Journal of International Financial Markets, Institutions and Money 22*(5), 1258–1276.

Amin, K. I. and A. J. Morton (1994). Implied volatility functions in arbitrage-free term structure models. *Journal of Financial Economics 35*(2), 141–180.

Ansotegui, C., A. Bassiouny, and E. Tooma (2013). The proof is in the pudding: Arbitrage is possible in limited emerging markets. *Journal of International Financial Markets, Institutions and Money 23*, 342–357.

Avellaneda, M. and J.-H. Lee (2010). Statistical arbitrage in the US equities market. *Quantitative Finance 10*(7), 761–782.

Backus, D., S. Foresi, and S. Zin (1998). Arbitrage opportunities in arbitrage-free models of bond pricing. *Journal of Business & Economic Statistics 16*(1), 13–26.

Black, F. (1976). The pricing of commodity contracts. *Journal of financial economics 3*(1–2), 167–179.

Bondarenko, O. (2003). Statistical arbitrage and securities prices. *Review of Financial Studies 16*(3), 875–919.

Bueno-Guerrero, A., M. Moreno, and J. F. Navas (2016). The stochastic string model as a unifying theory of the term structure of interest rates. *Physica A: Statistical Mechanics and its Applications 461*, 217–237.

Burgess, A. N. (1999). *A computational methodology for modelling the dynamics of statistical arbitrage.* Ph. D. thesis, University of London, London Business School.

Burgess, A. N. (2005). Using cointegration to hedge and trade international equities. In *Applied Quantitative Methods for Trading and Investment*, pp. 41–69. John Wiley & Sons, Ltd.

Choi, H., P. Mueller, and A. Vedolin (2017). Bond variance risk premiums. *Review of Finance 21*(3), 987–1022.

Dunis, C. and P. Lequeux (2000). Intraday data and hedging efficiency in interest spread trading. *The European Journal of Finance 6*(4), 332–352.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance 25*(2), 383–417.

Figlewski, S. (1989). Options arbitrage in imperfect markets. *The Journal of Finance 44*(5), 1289–1311.

Figlewski, S. (2016). Derivatives valuation based on arbitrage: The trade is crucial. *Journal of Futures Markets 37*(4), 316–327.

Filipović, D., M. Larsson, and A. B. Trolle (2017). Linear-rational term structure models. *The Journal of Finance 72*(2), 655–704.

Filipović, D. and A. B. Trolle (2013). The term structure of interbank risk. *Journal of Financial Economics 109*(3), 707–733.

Goldstein, R. S. (2000). The term structure of interest rates as a random field. *The Review of Financial Studies 13*(2), 365–384.

Harrison, J. M. and D. M. Kreps (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory 20*(3), 381–408.

Heath, D., R. Jarrow, and A. Morton (1990). Bond pricing and the term structure of interest rates: A discrete time approximation. *Journal of Financial and Quantitative analysis 25*(4), 419–440.

Heath, D., R. Jarrow, and A. Morton (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, 77–105.

Ho, T. S. Y. and S.-B. Lee (1986). Term structure movements and pricing interest rate contingent claims. *The Journal of Finance 41*(5), 1011–1029.

Hung, S.-Y., T.-P. Liang, and V. W.-C. Liu (1996). Integrating arbitrage pricing theory and artificial neural networks to support portfolio management. *Decision Support Systems 18*(3-4), 301–316.

Jagannathan, R., A. Kaplin, and S. Sun (2003). An evaluation of multi-factor CIR models using LIBOR, swap rates, and cap and swaption prices. *Journal of Econometrics 116*(1-2), 113–146.

Jarrow, R., H. Li, and F. Zhao (2007). Interest rate caps "smile" too! but can the libor market models capture the smile? *The Journal of Finance 62*(1), 345–382.

Krauss, C. (2016). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys 31*(2), 513–545.

Krauss, C., X. A. Do, and N. Huck (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research 259*(2), 689–702.

Liu, J., F. A. Longstaff, and R. E. Mandell (2006). The market price of risk in interest rate swaps: The roles of default and liquidity risks. *The Journal of Business 79*(5), 2337–2359.

Lo, A. W. (2004). The adaptive markets hypothesis. *Journal of Portfolio Management 30*(5), 15–29.

Lo, A. W. and A. C. MacKinlay (1988). Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies 1*(1), 41–66.

Lo, A. W. and J. Wang (1995). Implementing option pricing models when asset returns are predictable. *Journal of Finance 50*, 87–129.

Longstaff, F. A., P. Santa-Clara, and E. S. Schwartz (2001a). The relative valuation of caps and swaptions: Theory and empirical evidence. *The Journal of Finance 56*(6), 2067–2109.

Longstaff, F. A., P. Santa-Clara, and E. S. Schwartz (2001b). Throwing away a billion dollars: The cost of suboptimal exercise strategies in the swaptions market. *Journal of Financial Economics 62*(1), 39–66.

Merton, R. C. et al. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science 4*(1), 141–183.

Mitra, S., V. R. Chinthalapati, E. Clark, and F. McGroarty (2019). Stock-ADR arbitrage: Microstructure risk. *Journal of International Financial Markets, Institutions and Money 63*, 101132.

Ritchken, P. and K. Boenawan (1990). On arbitrage-free pricing of interest rate contingent claims. *The Journal of Finance 45*(1), 259–264.

Ross, S. A. (2005). *Neoclassical Finance.* Princenton University Press.

Santa-Clara, P. and D. Sornette (2001). The dynamics of the forward interest rate curve with stochastic string shocks. *The Review of Financial Studies 14*(1), 149–185.

United Nations (2019). National Accounts – Analysis of Main Aggregates (AMA). Technical report, Economic Statistics Branch of the United Nations Statistics Division. Retrieved May 20, 2020, from https://unstats.un.org/unsd/snaama/Index.

Urrutia, J. L. (1995). Tests of random walk and market efficiency for Latin American emerging equity markets. *Journal of Financial Research 18*(3), 299–309.

Xia, Y. (2001). Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation. *The Journal of Finance 56*(1), 205–246.

Table I Likelihood Ratio Test

The table reports results of the likelihood ratio test, that make pairwise comparisons between the models with N and N+1 factors. The difference between the sum of the squared errors is asymptotically χ^2_{61} under the null hypothesis of equality for the sample with 61 weeks and χ^2_{30} for the two half samples. The critical value of χ^2_{61} is 89.59 at the 99% level of confidence and for χ^2_{30} is 50.89 for the same level of confidence.

N Factors	$N+1$ Factors	Statistical Test	P-Value
	Panel A. Full Sample		
1	2	6,261,40	0.00
2	3	391.05	0.00
3	4	632.97	0.00
4	5	69.31	0.23
	Panel B. First Half of the Sample		
1	2	2,295.60	0.00
2	3	161.22	0.00
3	4	354.51	0.00
4	5	13.43	1.00
	Panel C. Second Half of the Sample		
1	2	3,565.70	0.00
2	3	225.59	0.00
3	4	268.18	0.00
4	5	53.52	0.01

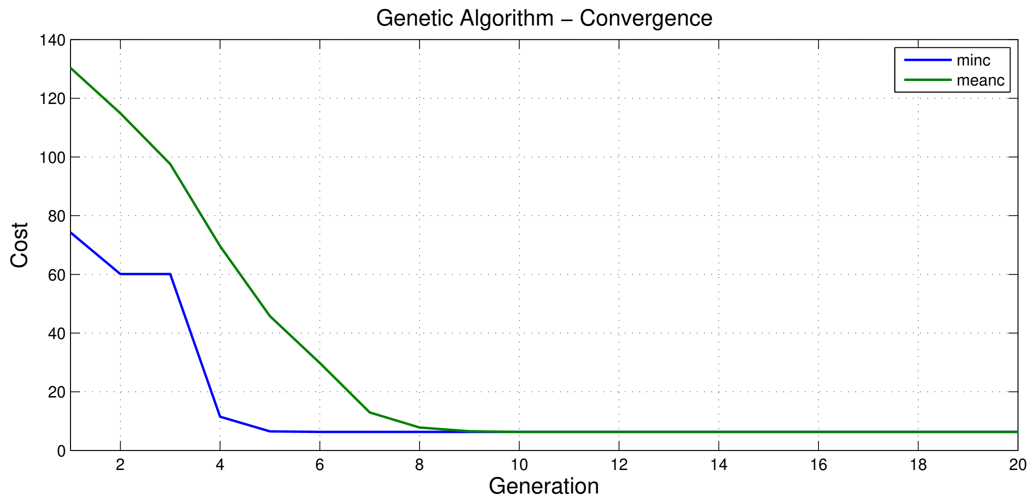

Figure 1. Genetic Algorithm Convergence. This figure shows the convergence of the genetic algorithm. Minc reflect the acting of the best member of the population in each interaction, Meanc is the arithmetic mean of the entire population. The y axis denotes the results of the objective function. To show the convergence, we begin with a stratospheric value of 500 to the volatility and the genetic algorithm finishes with 0.1.

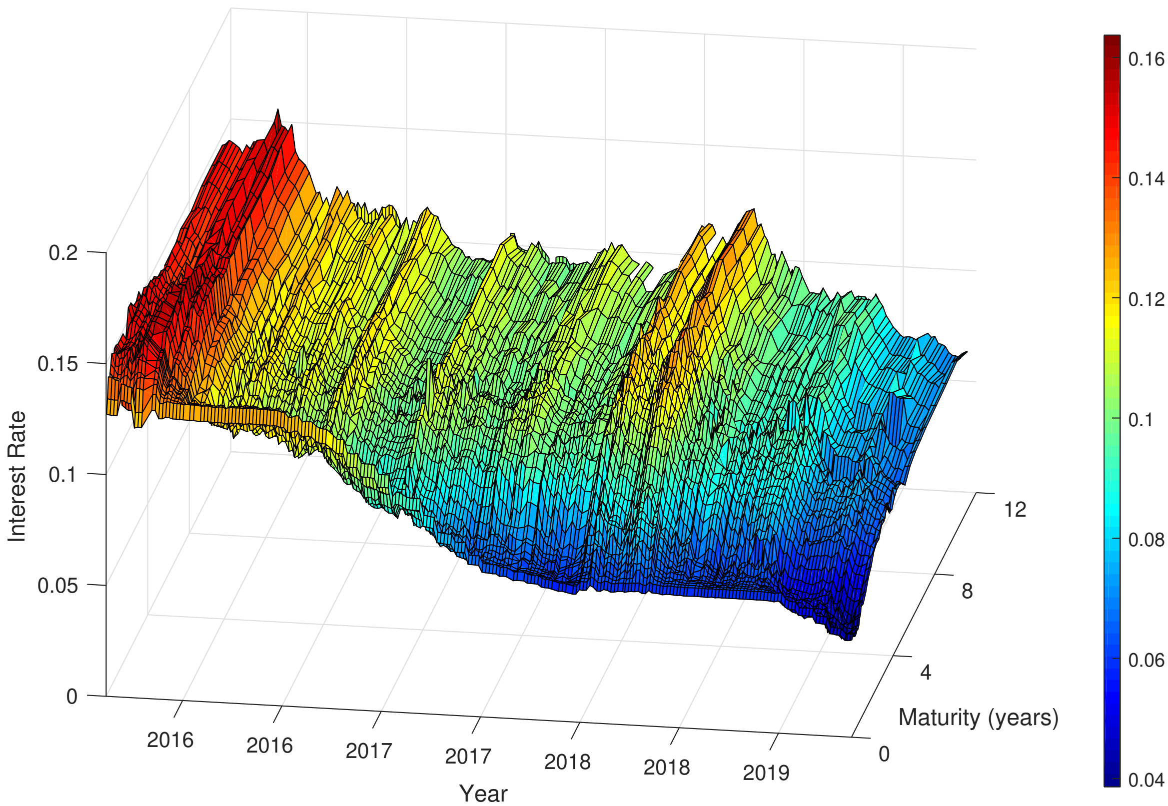

Figure 2. Evolution of the future interest rates – DI futures contract. This figures shows the DI futures term structure evolution. The data set consists of observations extracted weekly from October 2015 to January 2019, using Friday closing rates quotes. The sample contain data of 164 different days and for each day there are approximately 40 maturity dates. The rates was interpolated with the cubic spline.

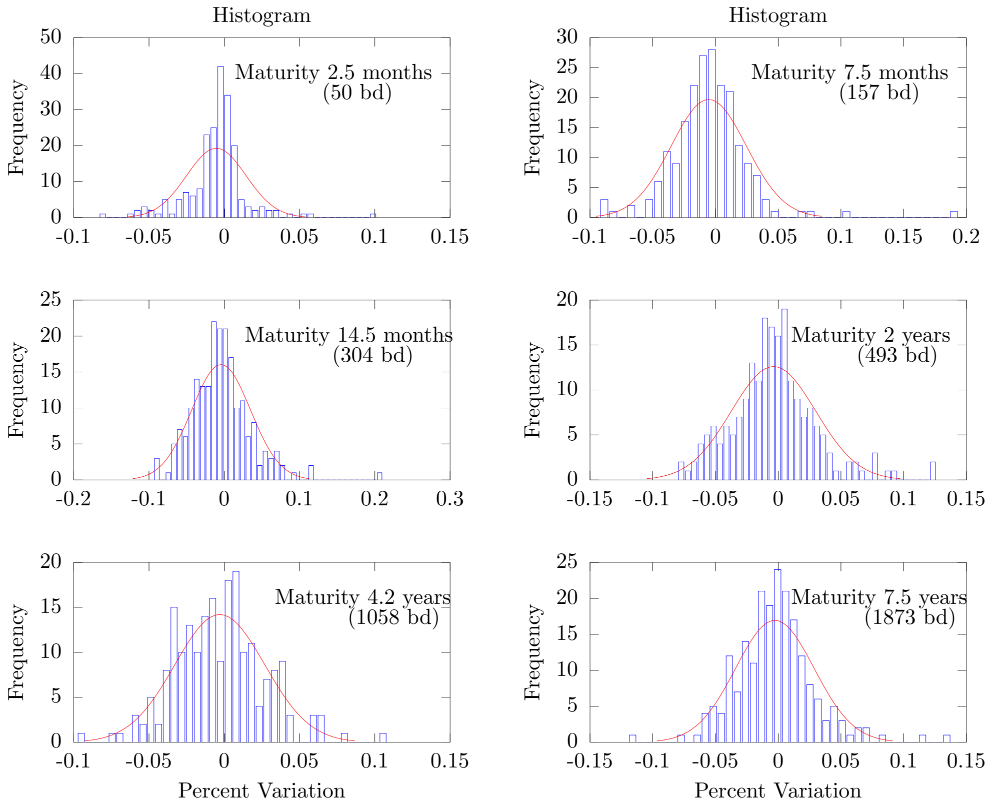

Figure 3. Histogram of the of the DI futures contract returns. DI futures returns are calculated using the annual rate with continuous capitalization. The graphs are divided based on the business day until the maturity of the contract, considering a year with 252 business day (bd).

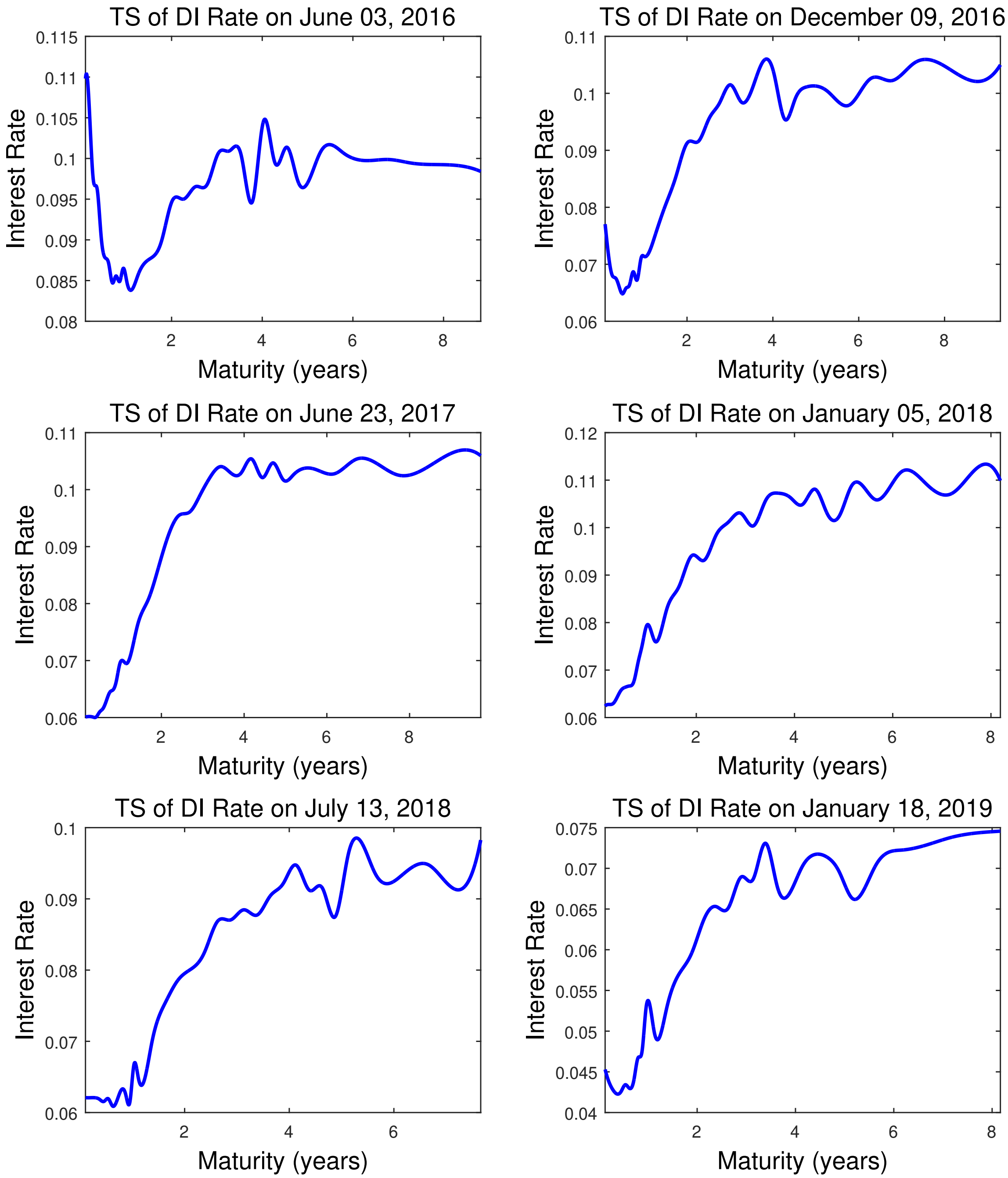

Figure 4. DI futures term structure curve. The future interest rates were interpolated using a cubic spline. The business day is on the horizontal axis and are obtained accordingly to the Brazilian bank calendar.

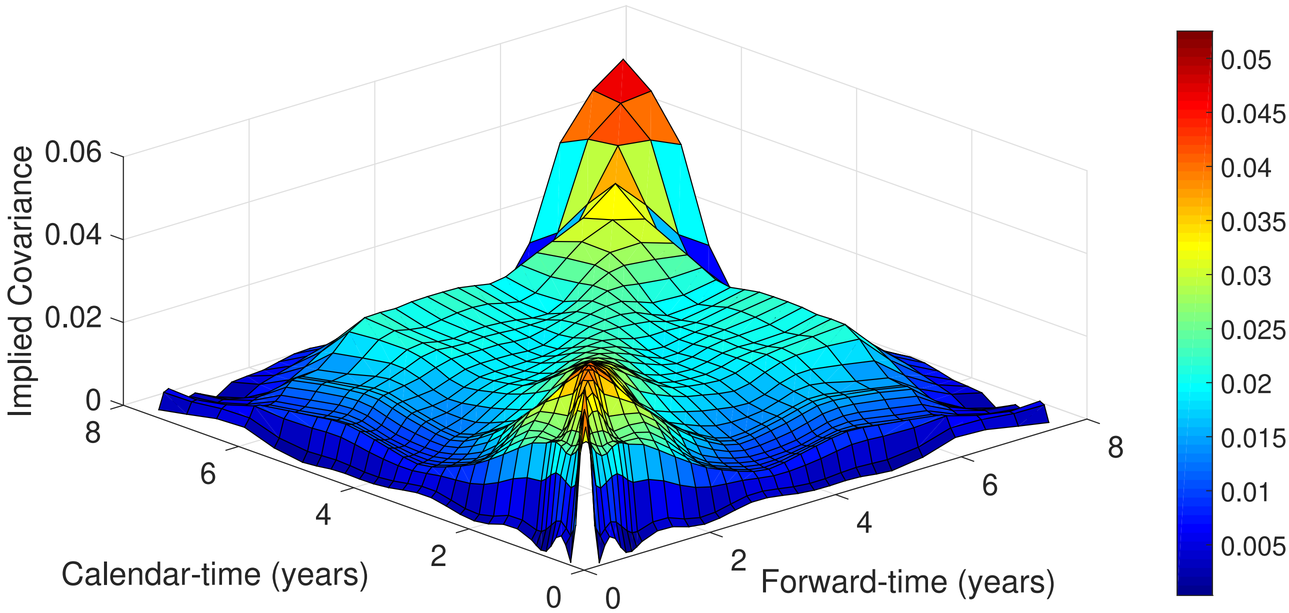

Figure 5. Implied covariance matrix. This figure presents the implied covariance matrix of the DI futures market, which was estimated with the implied volatility of the *DI future option* for January 18, 2019.

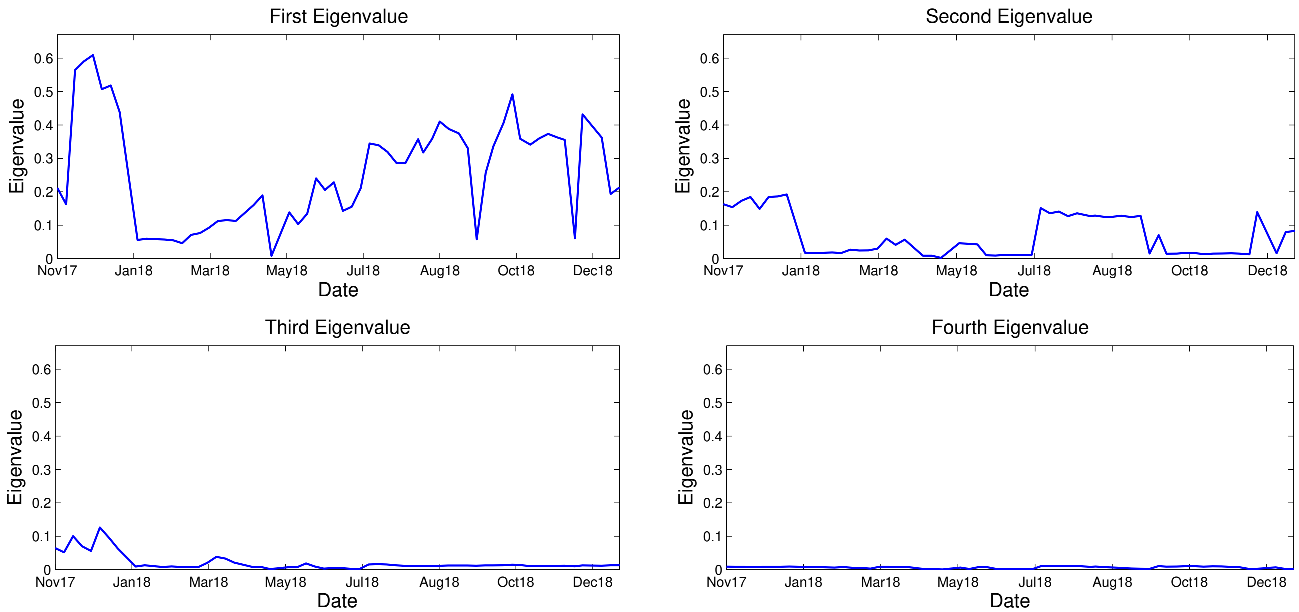

Figure 6. Implied Volatility Eigenvalues. This figure shows the time series of the eigenvalues related to the 61 weeks, from November 03, 2017 to January 18, 2019, all options of the sample are *at the money*. The eigenvalues are obtained from the implied volatility of the options on one-day interbank deposit futures.

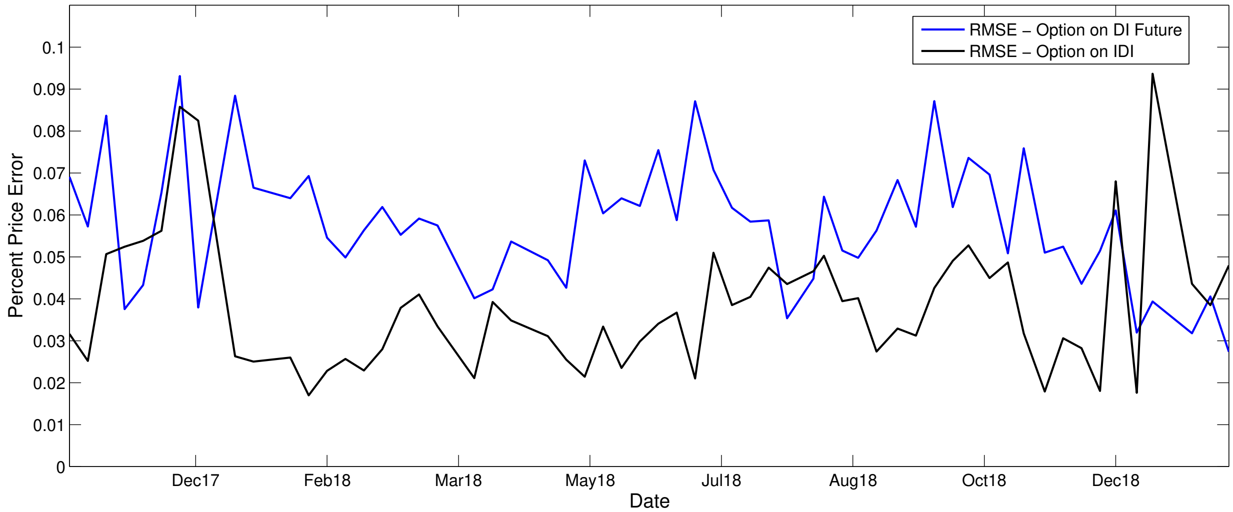

Figure 7. Statistical arbitrage in all In-the-Money (ITM) options. The time series of the RMSE of the *DI futures options*, and the *IDI index options*, from November 03, 2017 to January 18, 2019. All options of the sample are at the money. There are 36 active *DI futures options*, and 100 *IDI index options* for each date. Quotes are Friday closing prices. Model prices used a market data calibrated Longstaff et al. (2001a) string market model with four factors.

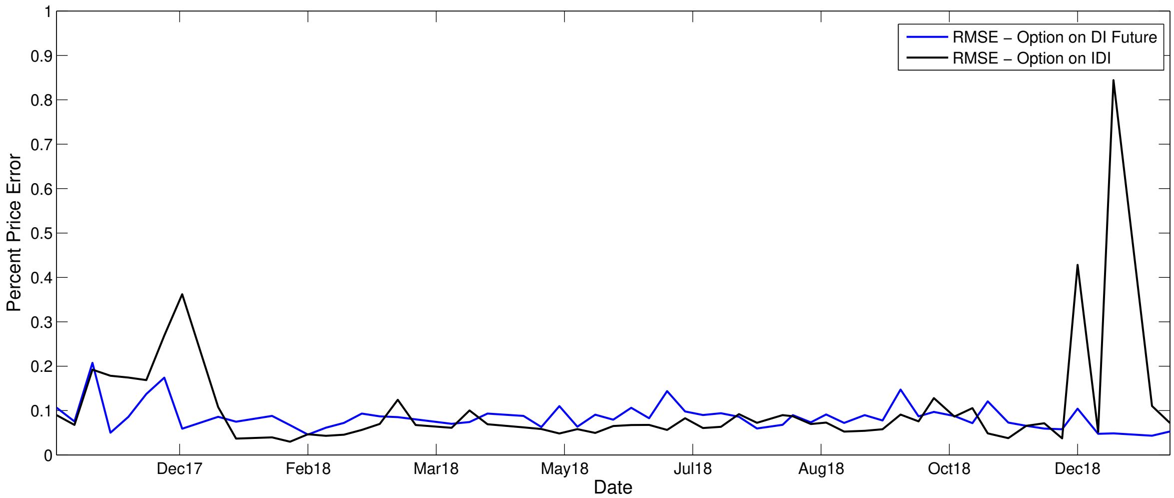

Figure 8. Statistical arbitrage in all Near-the-Money (NTM) options. The RMSE of the near-the-money (NTM) sub-sample of (i) the one-day interbank *DI futures options* (ii) and the one-day interbank deposit rate index (IDI) options. The sub-sample selects only near-the-money options from November 03, 2017 to January 18, 2019. Model prices used a market data calibrated Longstaff et al. (2001a) string market model with four factors.

Appendix A. Dynamic Evolution of the Bond Prices

Let $P(t,T)$ be the price at time t of a bond that has maturity at the date T, $P(T,T)=1$ and $P(t,T)>0$. The instantaneous interest rates at time t for a period between t and $T>0$, are denoted by $f(t,T)$. The spot interest rates $r(t)$, are the instantaneous interest rates available at time t, with time to maturity t, namely, $r(t)=f(t,t)$. Then,

$$\begin{aligned} f(t,T) &= -\log\left(\frac{\partial P(t,T)}{\partial T}\right), \\ P(t,T) &= \exp\left(-\int_0^T f(t,T)df\right), \\ r(t) &= f(t,t). \end{aligned}$$

The LSS dynamic of interest rates (we adopt for the DI rate, the same dynamic as the LIBOR rate), in the risk-neutral measure are,

$$\frac{dF(t,T_{j-1})}{F(t,T_{j-1)})} = \hat{\alpha}_{j-1}(t)dt + \bar{\sigma}_{j-1}(t)d\tilde{Z}_{j-1}(t), \tag{A1}$$

With,

$$\hat{\alpha}_{j-1}(t) = \frac{1+\delta F(t,T_{j-1})}{\delta F(t,T_{j-1})}\int_{y=T_{j-1}-t}^{T_j-t}\int_{u=0}^{T_j-t} R_t(u,y)dydu, \tag{A2}$$

$$\bar{\sigma}_{j-1}(t)d\tilde{Z}_{j-1}(t) = \int_{y=T_{j-1}-t}^{T_j-t} d\tilde{Z}(t,y)\bar{\sigma}(t,y)dy, \tag{A3}$$

$$R_t(u,y) = \frac{cov[df(t,u),df(t,y)]}{Ft} = c(t,x,y)\sigma(t,x)\sigma(t,y), \tag{A4}$$

where,

$$c(t,x,y) = \frac{d[Z(.,x),Z(.,y)]_t}{dt} = corr[d[Z(t,x),Z(t,y)],$$

LSS volatility:

$$\sigma_{i-1}^2(t) = \int_{x=T_{i-1}-t}^{T_i-t}\int_{y=T_{i-1}-t}^{T_i-t} R_t(x,y)dxdy. \tag{A5}$$

Historic Covariance Matrix:

$$\Theta_{ij}(t) = \int_{x=T_{i-1}-t}^{T_i-t}\int_{y=T_{j-1}-t}^{T_j-t} R(x,y)dxdy. \tag{A6}$$

Implied Covariance Matrix:

$$\Sigma_{ij}(t,T_0) = \int_{x=T_0-t}^{T_i-t}\int_{y=T_0-t}^{T_j-t} R(x,y)dxdy. \tag{A7}$$

Consider the interval $[T_0, T_n]$ and the partition $\{T_j = T_0 + \delta j\}_{j=1}^{n}$, where $\delta = \frac{T_n - T_0}{n}$, then:

$$P(T_{j-1}, T_j) = [1 + \delta F(T_{j-1})]^{-1}, \tag{A8}$$

and,

$$1 + \delta F(T_{j-1}) = \frac{P(t, T_{j-1})}{P(t, T_j)}. \tag{A9}$$

The bond prices evolve in accordance with:

$$dP(t) = r(t)P(t) + J^{-1}(t)\bar{\sigma}(t)F(t)d\tilde{Z}(t), \tag{A10}$$

where,

$$\begin{aligned} P(t) &= (P(t, T_1), ..., P(t, T_{n-1}))', & \text{(A11)} \\ \bar{\sigma}(t)F(t)d\tilde{Z}(t) &= (\bar{\sigma}_0 F(t, T_0)d\tilde{Z}_0(t), ..., \bar{\sigma}_{n-2}F(t, T_{n-2})d\tilde{Z}_{n-2}(t)), & \text{(A12)} \end{aligned}$$

and $J(t)$ is the Jacobian matrix, with $J_{ii}(t) = -\frac{1}{\delta}\frac{P(t,T_{i-1})}{P^2(t,T_i)}$ and $J_{i,i-1}(t) = \frac{1}{\delta P(t,T_i)}$, for $i = 1, ..., n-1$ and zero for the remainder terms. The introduction of the spot interest rates at expression (A10) is a manner of imposing the no-arbitrage condition. Let us apply this assumption at (A1) to verify that the expression (A10) is correct.

Consider that:

$$P(t, T_{j-1}) = \frac{P(t, T_0)}{\prod_{k=0}^{j-2}[1 + \delta F(t, T_k)]}, \quad j = 2, ..., n, \tag{A13}$$

So, applying the Itô's rule to $P(t, T_{j-1})$,

$$\begin{aligned} dP(t, T_{j-1}) = &\sum_{i=0}^{j-2} \frac{\partial P(t, T_{j-1})}{\partial F(t, T_i)} dF(t, T_i) + \frac{1}{2}\sum_{i,l=0}^{j-2} \frac{\partial^2 P(t, T_{j-1})}{\partial F(t, T_i), \partial F(t, T_l)} d[F(., T_i), F(., T_l)]_t) \\ &+ \frac{\partial P(t, T_{j-1})}{\partial P(t, T_0)} dP(t, T_0) + \sum_{i=0}^{j-2} \frac{\partial^2 P(t, T_{j-1})}{\partial F(t, T_i), \partial P(t, T_0)} d[F(., T_i), P(., T_0)]_t \\ &+ \frac{1}{2}\frac{\partial^2 P(t, T_{j-1})}{\partial^2 P(t, T_0} d[P(., T_0), P(., T_0)]_t. \end{aligned}$$

Expanding the first term of $dP(t, T_{j-1})$

$$\sum_{i=0}^{j-2} \frac{\partial P(t, T_{j-1})}{\partial F(t, T_i)} [\hat{\alpha}_i(t)F(t, T_i)dt + \bar{\sigma}_i(t)F(t, T_i)d\tilde{Z}_i(t)], \tag{A14}$$

where,

$$\frac{\partial P(t, T_{j-1})}{\partial F(t, T_i)} = \frac{-\delta P(t, T_{i+1})}{P(t, T_i)} P(t, T_{j-1}), \tag{A15}$$

Then,

$$\sum_{i=0}^{j-2} \frac{-\delta P(t,T_{i+1})}{P(t,T_i)} P(t,T_{j-1})[\hat{\alpha}_i(t)F(t,T_i)dt + \bar{\sigma}_i(t)F(t,T_i)d\tilde{Z}_i(t)]. \tag{A16}$$

Substituting the expression (A2) into the expression above,

$$\begin{aligned} \sum_{i=0}^{j-2} \frac{-\delta P(t,T_{i+1})}{P(t,T_i)} P(t,T_{j-1})F(t,T_i)\frac{1+\delta F(t,T_i)}{\delta F(t,T_i)} \int_{y=T_i-t}^{T_{i+1}-t} \int_{u=0}^{T_{i+1}-t} R_t(u,y)dydu \\ = -P(t,T_{j-1}) \sum_{i=0}^{j-2} \int_{y=T_i-t}^{T_{i+1}-t} \int_{u=0}^{T_{i+1}-t} R_t(u,y)dydu\, dt. \end{aligned} \tag{A17}$$

By the symmetry of $R_t(x,y)$, the expression (A17) is equal to,

$$-P(t,T_{j-1}) \int_{y=T_0-t}^{T_{j-1}-t} \int_{u=0}^{T_0-t} dyduR_t(u,y)dt. \tag{A18}$$

With relation to the second part of the equation (A14),

$$\sum_{i=0}^{j-2} \frac{-\delta P(t,T_{i+1})}{P(t,T_i)} P(t,T_{j-1})\bar{\sigma}_i(t)F(t,T_i)d\tilde{Z}_i(t). \tag{A19}$$

For the remaining terms, by introducing the equation (A3) in (A1),

$$\frac{dF(t,T_{j-1})}{F(t,T_{j-1}))} = \hat{\alpha}_{j-1}(t)dt + \int_{T_{j-1}-t}^{T_j-t} dy\bar{\sigma}(t,y)\tilde{Z}(t,y). \tag{A20}$$

Henceforth, we use the following rule: $dt^2 = 0$, $dt.dZ = 0$ and $dZ^2 = dt$.

Expanding the second term of $dP(t,T_{j-1})$

$$\begin{aligned} \frac{1}{2}\sum_{i,l=0}^{j-2} \frac{\partial^2 P(t,T_{j-1})}{\partial F(t,T_i,\partial F(t,T_l)} [\hat{\alpha}_i(t)F(t,T_i)dt + \bar{\sigma}_i(t)F(t,T_i)d\tilde{Z}_i(t), \\ \hat{\alpha}_l(t)F(t,T_l)dt + \bar{\sigma}_l(t)F(t,T_l)d\tilde{Z}_i(t)]. \end{aligned}$$

Applying the rules above:

$$\frac{1}{2}\sum_{i,l=0}^{j-2} \frac{\partial^2 P(t,T_{j-1})}{\partial F(t,T_i,\partial F(t,T_l)} \int_{T_{i-1}-t}^{T_i-t} dy\bar{\sigma}(t,y)\tilde{Z}(t,y) \int_{T_{l-1}-t}^{T_l-t} dy\bar{\sigma}(t,y)\tilde{Z}(t,y).$$

Once again, by the symmetry of $R_t(x,y)$, the term above reaches zero.

Expanding the third term of $dP(t,T_{j-1})$

Considering the follow dynamic of the bonds prices, according with Bueno-Guerrero et al. (2016),

$$\frac{dP(t,T_0)}{P(t,T_0)} = r(t)dt - \int_{y=0}^{T_0-t} dy\bar{\sigma}(t,y)\tilde{Z}(t,y), \tag{A21}$$

Then, with relation to the third term,

$$P(t,T_{j-1})\left(r(t)dt - \int_{y=0}^{T_0-t} dy\bar{\sigma}(t,y)\tilde{Z}(t,y)\right) \tag{A22}$$

Expanding the fourth term of $dP(t,T_{j-1})$

$$\begin{aligned}
&\sum_{i=0}^{j-2}\frac{\delta}{1+\delta F(t,T_i)}\frac{1}{\prod_{k=0}^{j-2}[1+\delta F(t,T_k)]}P(t,T_0)F(t,T_i)\int_{T_i-t}^{T_{i+1}-t} dy\bar{\sigma}(t,y)\tilde{Z}(t,y)\int_{y=0}^{T_0-t} dy\bar{\sigma}(t,y)\tilde{Z}(t,y) = \\
&\qquad = \sum_{i=0}^{j-2}\frac{\delta F(t,T_i)}{1+\delta F(t,T_i)}P(t,T_{j-1}).\int_{T_i-t}^{T_{i+1}-t} dy\bar{\sigma}(t,y)\tilde{Z}(t,y)\int_{y=0}^{T_0-t} dy\bar{\sigma}(t,y)\tilde{Z}(t,y) \\
&\qquad = P(t,T_{j-1})\int_{y=T_0-t}^{T_{j-1}-t}\int_{u=0}^{T_0-t} dyduR_t(u,y)dt.
\end{aligned} \tag{A23}$$

Expanding the fifth term of $dP(t,T_{j-1})$

Given that,

$$\frac{1}{2}\frac{\partial^2 P(t,T_{j-1})}{\partial^2 P(t,T_0)} = 0. \tag{A24}$$

Then, this term is dropped.

Aggregation of the terms

Adding the equations (A18), (A19), (A22), and (A23), and after that, dividing everything by $P(t,T_{j-1})$,

$$\frac{dP(t,T_{j-1})}{P(t,T_{j-1})} = r(t)dt - \sum_{i=0}^{j-2}\frac{-\delta P(t,T_{i+1})}{P(t,T_i)}\bar{\sigma}_i(t)F(t,T_i)d\tilde{Z}_i(t) - \int_{y=0}^{T_0-t} dy\bar{\sigma}(t,y)\tilde{Z}(t,y). \tag{A25}$$

In accordance with Longstaff et al. (2001a) (n. 12), for $y \leq (T_0-t)$ the volatility $\sigma(t,y)=0$, in other words, the process is not stochastic and do not affect the diffusion term in equation (A10). Therefore, the last term of the expression of (A25) is dropped. Take into account that:

$$[\mathbf{J}^{-1}(t)]_{ij} = \begin{cases} \frac{-\delta P(t,T_{i+1})}{P(t,T_i)} & if\ j \leq i \\ 0 & if\ j > i, \end{cases}$$

we get,

$$dP(t) = r(t)P(t) + J^{-1}(t)\bar{\sigma}(t)F(t)d\tilde{Z}(t).$$