Department of Economics, Finance & Accounting

Implicit Entropic Market Risk-Premium from Interest Rate Derivatives[*]

J. Arismendi-Zambrano[b,1,*], R. Azevedo[c]

[a]*Department of Economics, Finance and Accounting, National University of Ireland, Maynooth, Ireland.*
[b]*ICMA Centre, Henley Business School, University of Reading, Whiteknights, Reading, United Kingdom.*
[c]*Bradesco Asset Management - BRAM, Sao Paulo, Brazil.*

Abstract

Implicit in interest rate derivatives are Arrow–Debreu prices (or state price densities, SPDs) that contain fundamental information for risk and portfolio management in interest rate markets. To extract such information from interest rate derivatives, we propose a nonparametric method to estimate state prices based on the minimization of the Cressie–Read (Entropic) family function between potential SPDs and the empirical probability measure. An empirical application of the method, in the US interest rates and derivatives market, shows that the *entropic based* risk-neutral density measure highlight potential risks previous to the 2007/2008 financial crisis, and the potential arbitrage burden during the *Quantitative Easing* period.

Keywords: Risk management, Risk analysis, Nonparametric Asset Pricing, State Price Density, Interest Rate Derivatives

JEL Classification: C14, G12, G13, G14, G18.

1. Introduction

Interest rate derivatives securities, such as caps, floors and swaptions, are among the most widely traded derivatives instruments in the world;[1] they embed market expectations on the volatility, and other higher-order moments of interest rates; for this reason they have been extensively used in economic analysis (See for instance, Li and Zhao, 2006, Bikbov and Chernov, 2013, and Almeida et al., 2011). In the valuation of assets, an assessment of the market expectations and uncertainty is needed; usually, financial risk-associated measures (market, credit, liquidity, operational) are defined and used to assess the amount of uncertainty in financial markets.

One of the recently defined group of risk measures for financial applications are the entropic: Frittelli (2000) and Rouge and Karoui (2000) used minimal entropy martingale measures for option pricing; Ahmadi-Javid (2011) defined a coherent Value-at-Risk (VaR) measure based in entropic risk, the *entropic VaR*, and then Ahmadi-Javid and Fallah-Tafti (2019) used the *entropic VaR* to formulate and solve the classical optimal portfolio/asset allocation problem, substituting the variance by the entropic risk measure; Brandtner et al.

(2018) discussed the *coherent entropic risk measures (CERM) and convex entropic risk measures (ERM)* first introduced by Föllmer and Knispel (2011) within the optimal portfolio allocation problem; and Bekiros et al. (2017) used a entropic dependence measure to analyze the contagion and network transmission effects between equity and commodity markets. A detailed survey of the application of entropy measures in finance can be found in Zhou et al. (2013).[2] and a rigorous mathematical analysis of the entropic risk measures in Pichler and Schlotter (2019).

In this paper we provide a framework to derive the entropic based risk-neutral density of the interest rate derivatives markets, in line with the *canonical valuation* option pricing method of Stutzer (1996).[3] Our study builds over two strands of the literature: (i) a theoretical, by extending the use of entropic risk measures in option pricing, as in Stutzer (1996), Rouge and Karoui (2000), Frittelli (2000), and Gzyl and Mayoral (2012), and (ii) an empirical, by extending the market implied risk-neutral density extraction in line with Aït-Sahalia and Lo (1998), Chernov and Ghysels (2000), Unal et al. (2003), Rompolis and Tzavalis (2008), Rompolis (2010), Li and Zhao (2009), and Fabozzi et al. (2009).

The contributions of this paper are threefold: First, this research provides a generalization of the *canonical valuation* method developed in Stutzer (1996) to interest rate derivatives. Although Stutzer and Chowdhury (1999) considered an extension of *canonical valuation* to bond futures options, they only consider the last maturity of the interest rate term structure, while we are considering an arbitrary M number of maturities of the interest rate term structure, that can be used to model the whole curve. Essentially, the *canonical valuation* defines the risk-neutral measure as the one which is closest, by an entropic distance measure, to the empirical distribution of the returns, and at the same time satisfies a set of restrictions based on pricing equations. The simplest set of restriction consists by the equation that prices the return itself (i.e. $E^{\mathbb{Q}}[R] = (1 + r^f)$) where the superscript $\mathbb{Q}$ refers to the risk-neutral measure). More restrictions can be added to price correctly the available data.

The *canonical valuation* method was applied in Stutzer (1996) using historical market prices and in a simulated Black and Scholes (1973) world; results show that the *canonical valuation* slightly underperforms the historical Black and Scholes (1973) price, but this difference is narrow considering that *canonical valuation* has no information about the underlying process (nonparametric method). Stutzer's (1996) results show that *canonical valuation* option pricing can model the volatility smile. Our *canonical valuation* option pricing for interest rate derivatives is a more general version of Stutzer (1996) as we have an inherently multidimensional setting that can handle more complex information structures such as the interest rate term structure. In our development, we priced a *cap*, but the method can be immediately applied to a greater set of derivatives securities.[4] Stutzer (1996) models future scenarios by using a grid with the historical data prices. In our approach, we used (i) the grid of historical prices from Stutzer (1996), but additionally we introduced the use of (ii) a grid with a greater range of scenarios, to model all the possible future scenarios of interest rate term structure evolution, named the plain grid; (iii) A third grid named the *Svensson (1994) grid*, similar to the *plain grid*, is developed, but including additional arbitrage constraints based on the Svensson (1994) interest rate term structure approximation.

---

[2]Entropic measures have been used in management and operational research applications before, see for instance: Hoskisson et al. (1993) for an application of entropy to diversification of companies' strategies, Shuiabi et al. (2005) for the measurement of a company's operational flexibility, and Fleischhacker and Fok (2015) to measure the product demand uncertainty

[3]The word "canonical" refers to the *Gibbs canonical distribution*, an essential distribution in the theory of statistical mechanics.

[4]We consider the possibility of incorporating different interest rate instruments maturities into the modeling at the same time, that is common characteristic of most interest rate derivative such as *caps*.

Second, the concept of implicit *entropic risk-neutral density premium* is introduced, to define the difference between market option prices and *canonical valuation* option prices. This contribution is related to papers trying to estimate the forward or risk-neutral measures (allowing giving prices to interest rate derivatives) through a nonparametric estimation. Our method is comparable to the Li and Zhao (2009) interest rates derivatives risk-neutral density by the nonparametric nature, but differs in the approach and the result by using entropic functions to measure the distance between the physical and the risk-neutral densities, instead of polynomial functions (Li and Zhao, 2009). (Li and Zhao, 2009) extended to interest rates derivatives, the Aït-Sahalia and Lo (1998) nonparametric kernel data based method for equity option pricing. In particular, what we do is related to the minimum discrepancy estimator as defined in Kitamura (2006). In other words, we assume that the empirical distribution is given and consider the set of all forward-measure distributions consistent with the data available on prices. Inside this set, we choose the forward-measure distribution that is closest to the empirical one. We use more than one notion of distance that is given by a generalization of the cross-entropy.[5] This leads to an unfeasible optimization problem with a very high dimensionality and we overcome this issue by establishing the dual problem. An interesting result is that this dual problem is equivalent to a portfolio problem with fixed-income assets.

Third, we provide a numerical and an empirical application, the numerical with an analysis of our method in the context of the Heath et al. (1992) framework (HJM), and the empirical by measuring the *entropic risk-neutral density* premium of US interest rates caps. In the numerical case we show that the method provides the right forward-measure distribution under the Heath et al. (1992) framework given sufficient conditions on the risk premium. As a particular case, we discuss the market LIBOR[6] model and show that the method provides the same price for caplets as the Black caplet formula. In the empirical case, by using a selection of US interest rate swaps, swaptions, and cap market prices from May 2005 to August 2013,[7] we measure the implicit *entropic risk-neutral density premium* of the interest rate derivatives (caps). In the empirical analysis, we find that the *entropic risk-neutral density premium* of the interest rates caps increases and peaks in March 2008, during the Bear Stearns default, and again in October 2008 during the Lehman Brothers default, but is reduced to almost zero by 2013, mainly due to the Quantitative Easing and other monetary policy actions that stabilized changes in the interest rate term structure. Our empirical results extend those of Rompolis (2010), and Rompolis and Tzavalis (2010) from the equity market to the interest rate markets. Similarly, we contribute to the literature on risk-neutral density properties, by partially extending the results from equity markets (Bali and Murray, 2013; Leiss and Nax, 2018; Barletta et al., 2019) to the interest rate markets.

This paper is organized as follows: Section 2 briefly introduces the *canonical valuation* notation as in Stutzer (1996). Section 3 describes our proposal for pricing fixed-income derivative. Section 4 discusses the method in the context of HJM framework. Section 5 presents the empirical application with the US interest rates derivatives, and Section 6 concludes.

---

[5]This generalization encompasses some well known divergence criteria such as Kulback–Leibler and Empirical Likelihood.

[6]*LIBOR* is an acronym of *London InterBank Offered Rate*, that represents the average interest rate at which the banks in London will lend between them in American dollars. US LIBOR will be equivalent in being an interbank rate, but for the US banks.

[7]The data window is defined to include the crisis period of 2007/2008 financial crisis; and it's limited to 8 years due to the computational complexity required to estimate the risk-neutral density with the *plain grid* and the Svensson (1994) grid methods

2. Implicit entropic risk measures: *canonical valuation*

Let $(\Omega, \mathcal{F}_t, \mathbb{Q})$ be a risk-neutral probability space, and consider pricing a European call option in time t with an expiration in $T = t + \tau$ and strike price K with a constant risk-free rate r (daily compounded). The option price in the risk-neutral measure Q is

$$C = E_t^{\mathbb{Q}} \left[\frac{\max\{P_T - K, 0\}}{(1+r)^\tau} \Big| \mathcal{F}_t \right], \tag{1}$$

where $E_t^{\mathbb{Q}}[.]$ indicates expectation in risk-neutral measure $\mathbb{Q}$ conditional on information $\mathcal{F}_t$, P_T is the price of underlying asset in date T. Suppose we have a discrete time series of underlying asset prices $\{P_t\}_{t=1}^T$ with no dividends/additional cash-flows, and define $\{R_t = P_T/P_t\}_{t=1}^T$ the asset's gross return. Stutzer (1996) considers as an estimation of the probability in the physical world measure $\mathbb{P}$, the *empirical distribution* that assigns to each observed return the same probability: $\pi_t^{\mathbb{P}} = 1/\tau$. Let Ω have N scenarios, this *entropic physical measure* $\pi^{\mathbb{P}}$ can be transformed into a *entropic risk-neutral* measure $\pi^{\mathbb{Q}}$ by satisfying in the absence of arbitrage

$$\sum_{k=1}^N \pi_k^{\mathbb{Q}} = 1, \qquad \sum_{k=1}^N \frac{R_k}{(1+r)^\tau} \pi_k^{\mathbb{Q}} = 1, \qquad \text{and,} \qquad \pi_k^{\mathbb{Q}} > 0, \forall k \in \{1, \ldots, N\}, \tag{2}$$

where $\pi_k^{\mathbb{Q}}$ is the probability in the risk-neutral measure, associated to R_k. In the case we consider a one-to-one correspondence between any observation of the time series $t \in \{1 \ldots, T\}$ and the possible scenarios $k \in \{1, \ldots, N\}$, then $k = t$, and $N = T$. In this case, we can price the call option in the entropic risk-neutral measure as usual

$$C_{\text{entropic}} = E_t^{\mathbb{Q}} \left[\sum_{t=1}^T \frac{\max\{P_t R_t - K, 0\}}{(1+r)^\tau} \pi_t^{\mathbb{Q}} \right]. \tag{3}$$

There are an infinite number of probability measures that satisfy the equations in (2). In selecting a risk-neutral measure, Stutzer (1996) considered the one that reduced the distance between $\pi^{\mathbb{P}}$ and $\pi^{\mathbb{Q}}$ (the *empirical distance*), but the definition of distance is not unique, for this reason, Stutzer (1996) selected the Kulback–Leibler Information Criterion (KLIC) distance[8] to find the risk-neutral measure closest to the empirical. As we have that $\pi_t^{\mathbb{P}} = 1/T$

$$I\left(\pi_t^{\mathbb{Q}}, \pi_t^{\mathbb{P}}\right) = \sum_{t=1}^T \pi_t^{\mathbb{Q}} \log\left(\frac{\pi_t^{\mathbb{Q}}}{\pi_t^{\mathbb{P}}}\right) = \sum_{t=1}^T \pi_t^{\mathbb{Q}} \log \pi_t^{\mathbb{Q}} - \log(T). \tag{4}$$

A risk-neutral estimation procedure reduces to the optimization problem

$$\pi_t^{\mathbb{Q}} = \arg\min_{\widetilde{\pi}_t^{\mathbb{Q}}} I\left(\widetilde{\pi}_t^{\mathbb{Q}}, \pi_t^{\mathbb{P}}\right), \tag{5}$$

$$s.t.$$

---

[8]More precisely, the KLIC is a divergence and not a distance. In particular, the KLIC is not symmetric, i.e., we do not have necessarily that $I\left(\pi^1, \pi^2\right) \neq I\left(\pi^2, \pi^1\right)$.

$$\sum_{t=1}^{T} \pi_t^{\mathbb{Q}} = 1, \tag{6}$$

$$\sum_{t=1}^{T} \frac{R_t}{(1+r)^T} \pi_t^{\mathbb{Q}} = 1, \tag{7}$$

$$\pi_t^{\mathbb{Q}} > 0, \tag{8}$$

Realize that in this case, minimizing $I\left(\pi_t^{\mathbb{Q}}, \pi_t\right)$ is the same as maximizing $-\sum_{t=1}^{T} \pi_t^{\mathbb{Q}} \log \pi_t^{\mathbb{Q}}$ which is the definition of the Shanon entropy, a fundamental quantity in information theory.[9]

2.1. Implicit entropic measure and the market prices - economic significance

Classical and entropic option prices (Eq. (1) and Eq. (3)) have the same value when $\pi_t^{\mathbb{P}}$ and all the N possible future scenarios are known, or can be correctly estimated; for example, when the underlying asset price process is binomial. Nevertheless, in empirical applications, the future scenarios are unknown and we have to produce an estimation of the *empirical distribution* in $\pi_t^{\mathbb{P}}$.

Given that market prices are available, we can invert the results of the *entropic risk-neutral measure* optimization problem (5), and use the call price Eq. (1) and Eq. (3) to extract the *implicit entropic risk measures* given by the market, in a similar manner to what Black (1976) does when extracting the implied volatility from the market prices.

Define C_m as the European call option market price. Let $\{P_k\}_{k=1}^{N}$ represent an estimated set of the N future scenarios. The difference between C_m and C_{entropic} is defined as the *implicit entropic price call-premium*

$$X_e = C_{\text{entropic}} - C_m, \tag{9}$$

and the difference between the physical measure and the estimated *entropic risk-neutral measure* is defined as the *entropic risk-neutral density premium*. In interest rate markets, the *implicit entropic price call-premium* (Eq. (9)) will have an economic meaning: the difference between the *entropic call prices* and market call prices represents the difference between (i) the risk-neutral *entropic* expectations, deduced from the market expectations of the physical measure imprinted in the interest rate term structure, and (ii) the market risk-neutral expectations.

3. Informational pricing of interest rate derivatives securities

We are interested in pricing interest rate derivatives securities. We will focus in pricing a cap but the procedure described here can be easily generalized for a wider class of interest rate derivatives securities. On

---

[9]It is useful to solve this problem in a different way. To this end, consider the Lagrangian

$$L = \sum_{t=1}^{N} \left\{ \pi_t^{\mathbb{Q}} \log \pi_t^{\mathbb{Q}} - \log(T) + \gamma_1 \left(\frac{R_t}{(1+r)^T} \pi_t^{\mathbb{Q}} - \frac{1}{T} \right) + \lambda \left(\pi_t^{\mathbb{Q}} - \frac{1}{T} \right) \right\}.$$

Working on it we obtain

$$\pi_t^{\mathbb{Q}} = \frac{\exp\left(\gamma_1 \frac{R_t}{(1+r)^h}\right)}{\sum_{k=1}^{N} \exp\left(\gamma_1 \frac{R_k}{(1+r)^h}\right)},$$

that following Ben-Tal (1985), corresponds to the Gibbs distribution, that plays a fundamental role in Statistical Mechanics; and

$$\gamma_1 = \arg\min_{\gamma} \sum_{t=1}^{T} \exp\left(\gamma \frac{R_t}{(1+r)^T} - \gamma\right).$$

We transformed a multidimensional optimization problem into a one-dimensional.

this class of problems, it is usually more convenient to use the forward measure. In $\mathbb{T}$−forward measure $Q^{\mathbb{T}}$, the price of the payoff H_T paid in T is

$$p(H_T) = P(t,T)E_t^{\mathbb{T}}[H_T],\tag{10}$$

where $E_t^{\mathbb{T}}[\cdot]$ denotes the expectation in the $Q^{\mathbb{T}}$ conditional to information set $\mathcal{F}_t$, and $P(t,T)$ is the zero-coupon bond price in t with maturity in T. For every maturity T_i, there is a corresponding $\mathbb{T}_i$ forward measure $E^{\mathbb{T}_i}$.

In order to fix the notation, we provide some definitions and standard results before defining the method.

3.1. Definitions and notation

In order to introduce the notation, we give some definitions in this section.[10] The simply compounded spot rate at time t for the maturity T is denoted by $L(t,T)$ and is related to the zero-coupon bond price by

$$L(t,T) = \frac{1 - P(t,T)}{\tau(t,T)P(t,T)},\tag{11}$$

where $\tau(t,T)$ is the day-count convention between the dates t and T. This means that if you borrow a nominal amount of Y at t and pay it in T at the rate $L(t,T)$, you should pay at T the amount $Y(1 + \tau(t,T)L(t,T))$.

The simply compounded forward rate at time t for the expiry–maturity pairs T_1, T_2 is denoted by $F(t; T_1, T_2)$ and is related to $P(t,T)$ by the formula

$$F(t; T_1, T_2) = \frac{1}{\tau(T_1, T_2)}\left(\frac{P(t,T_1)}{P(t,T_2)} - 1\right).\tag{12}$$

The cap is a derivative defined for a payment flow depending upon a strike K. Consider a payment flow for dates $T_1, \ldots, T_M$ whose values are defined by some fluctuating rating (as the LIBOR rate) in $T_0, \ldots, T_{M-1}$ over the nominal amount Y (i.e. the payment in T_1 is known in T_0, the payment in T_2 is known in T_1 and so on). Let $\tau_i = \tau(T_{i-1}, T_i)$ to simplify the notation. So, in T_i the cap pays $= Y\tau_i\left(L(T_{i-1}, T_i) - K\right)^+$.

In the *canonical valuation* it was assumed that the risk-free rate did not change through time. This often proves to be a useful approximation when dealing with equities but this is no longer the case for fixed-income derivatives. Moreover it is usual to use the $\mathbb{T}$-forward measure and in this measure the restrictions above are no longer true. From now on we will make use of forward measures.

In order to obtain the $\mathbb{T}$-forward measure choose the zero coupon bond maturing at T as a numeraire. Let V_t be the contingent claim asset, then the asset prices described in terms of this numeraire can be written as martingales in the $\mathbb{T}$-forward measure

$$\frac{V_t}{P(t,T)} = E^{\mathbb{T}}\left[\frac{V_T}{P(T,T)}\Big|\mathcal{F}_t\right].\tag{13}$$

If the asset pays some time S before T, the above equation is rewritten as

$$V_t = P(t,T)E^{\mathbb{T}}\left[\frac{V_S}{P(S,T)}\Big|\mathcal{F}_t\right].\tag{14}$$

The above equation can be interpreted as a dynamical portfolio that pays $\frac{V_S}{P(S,T)}$ at T. In order to see this,

---

[10]See Chapter 2 of Brigo and Mercurio (2006) and Chapter 9 of Shreve (2004).

interpret $\frac{V_S}{P(S,T)}$ at S as the number of zero coupon bonds maturing at T invested in S with the money paid by the asset.

The above identities imply that the simply compounded forward rates are martingales. A way to see this is to rewrite the definition as

$$F(u;S,T) \quad = \quad \frac{1}{\tau(S,T)}\left(\frac{P(u,S)}{P(u,T)}-1\right)=\frac{\frac{1}{\tau(S,T)}\left(P(u,S)-P(u,T)\right)}{P(u,T)}, \tag{15}$$

and consider the right hand side as the portfolio's value in terms of the numeraire. Note that the portfolio considered is a long position of $\frac{1}{\tau(S,T)}$ bonds maturing in S and a short position of $\frac{1}{\tau(S,T)}$ maturing in T. With this in mind, it becomes clear that $F(u;T_1,T_2)$ is a martingale in the $\mathbb{T}_2$-forward measure

$$F(u;T_1,T_2) = E^{\mathbb{T}}\left[F(t;T_1,T_2)|\mathcal{F}_u\right], \tag{16}$$

and as a particular case we have

$$E^{\mathbb{T}_2}\left[L(T_1,T_2)|\mathcal{F}_u\right] = F(u;T_1,T_2), \tag{17}$$

when $t = S$ because $F(S;S,T) = L(T_1,T_2)$.

Finally the price of a cap in $\mathbb{T}_M-$forward measure is

$$p_t^{\mathrm{cap}} = P(t,T_M)E^{\mathbb{T}_M}\left[\sum_{i=1}^{M}\frac{Y\tau_i\left(L(T_{i-1},T_i)-K\right)^+}{P(T_i,T_M)}\bigg|\mathcal{F}_t\right]. \tag{18}$$

It is useful to consider a cap as being a set of caplets and the equation above simplifies to

$$p_t^{\mathrm{cap}} = \sum_{i=1}^{M} p_t^{\mathrm{caplet}_i}. \tag{19}$$

3.2. The random variables relevant for the method

We want to generalize the *canonical valuation* of European calls to caps. In brief, *canonical valuation* estimates the distribution of the price of underlying assets in a real world measure and then finds the closest distribution to this estimate that satisfies some identities. Those identities ensure that the latter distribution is a risk-neutral measure consistent with the available data. In the present work, we will find a $\mathbb{T}_M-$forward measure "closest" to the empirical distribution and consistent with the available data. In order to avoid theoretical measure considerations, we will consider that the empirical distribution is given by a discrete distribution π.

First it is important to define what the relevant random variables are. For the equities derivatives, the relevant one is the equity price and there is no need to model interest rates if the zero coupon bond price is available and we have only one random variable needed: the underlying asset. On the other hand, most interest rate derivatives deal with multidimensional random variables. For instance, caps can be separated into a stream of caplets and each caplet depends on a different spot rate. Moreover, it is necessary to have more random variables if someone wants to use the same forward measure to price all of them. In order to be more specific,

the formula to price the i-th caplet in the $\mathbb{T}_M-$forward measure is

$$p_t^{\text{caplet}_i} = P(t, T_M) E^{\mathbb{T}_M} \left[\frac{Y \tau_i \left(L(T_{i-1}, T_i) - K \right)^+}{P(T_i, T_M)} \bigg| \mathcal{F}_t \right]. \tag{20}$$

Eq. (20) shows that we need to find the distribution of $P(T_i, T_M)$ and $P(T_{i-1}, T_i)$ if $T_i \neq T_M$. For the case $T_i = T_M$, the equations above becomes

$$p_t^{\text{caplet}_M} = P(t, T_M) E^{\mathbb{T}_M} \left[Y \tau_i \left(L(T_{M-1}, T_M) - K \right)^+ | \mathcal{F}_t \right]. \tag{21}$$

Summing up, we need so far to consider the distribution of: $P(T_{M-1}, T_M)$ for $i = 1, 2, \ldots, M$ and $P(T_i, T_M)$ for $i = 0, 1, \ldots, M - 1$. There are still more relevant random variables because of the mismatch between the spot rate and the maturity of the bond used as the numeraire in the $\mathbb{T}_M-$forward measure. Before considering those we will define the method for the last caplet, as it is much simpler.

3.3. The method for the last caplet

For the last caplet, we have that $T_i = T_M$ and $P(T_i, T_M) = P(T_i, T_M) = 1$ and its price can be written as

$$p_t^{\text{caplet}_M} = P(t, T_M) E^{\mathbb{T}_M} \left[Y \tau_{M-1} \left(L(T_{M-1}, T_M) - K \right)^+ | \mathcal{F}_t \right]. \tag{22}$$

The only relevant random variable in this case is $L(T_{M-1}, T_M)$ or, equivalently, $P(T_{M-1}, T_M)$. In the $\mathbb{T}_M-$forward measure, $F(t; T_{M-1}, T_M)$ is a martingale and we have that

$$F(t; T_{M-1}, T_M) = E^{\mathbb{T}_M} \left[L(T_{M-1}, T_M) | \mathcal{F}_t \right]. \tag{23}$$

Approximating the empirical distribution of $L(T_{M-1}, T_M)$ by a discrete one with N states and denoting by π_k the empirical probability of the k-th state, our problem is to find the distribution $\overline{\pi}_k$ in the $\mathbb{T}_M-$forward measure that solves the following minimization problem

$$\overline{\pi}_k = \arg\min_{\overline{\pi}_k} I\left(\overline{\pi}_k, \pi_t \right), \tag{24}$$

$$s.t.$$

$$\sum_{k=1}^{N} \overline{\pi}_k = 1, \tag{25}$$

$$\overline{\pi}_k > 0 \text{ for } k = 1, 2, \ldots, N, \tag{26}$$

$$\sum_{k=1}^{N} \overline{\pi}_k \left(L_k(M - 1, M) \right) = F(t; T_{M-1}, T_M), \tag{27}$$

where $L_k(M - 1, M)$ is the simply compounded spot rate value $L(T_{M-1}, T_M)$ for the $k-$th state and the forward rate $F(t; T_{M-1}, T_M)$ is observed at t.

Now, to price the last caplet it is necessary only to apply the expression

$$p_t^{\text{caplet}_M} = P(t, T_M) \sum_{k=1}^{N} \overline{\pi}_k \left[Y \tau_{M-1} \left(L(T_{M-1}, T_M) - K \right)^+ \right]. \tag{28}$$

It is possible to price the $i-$th caplet in the same way using the $\mathbb{T}_i-$forward measure. It is a simple way to

generalize the method introduced by Stutzer (1996).

On the other hand, it is possible to price the $i-$th caplet using the $\mathbb{T}_M-$forward measure, but the method needs to accommodate more relations. This is interesting as the observed yield curve provides much more information that can be used in characterizing the $T_{\mathbb{M}}-$forward measure.

3.4. Equations used as restrictions

From now on we will use the $\mathbb{T}_M-$forward measure although most caplets pay at times T_i before T_M. In this measure we need to find equations similar to Eq. (27). Moreover, it will be interesting to express those equations using only bond prices as well. This is accomplished in the next result.

Proposition 1. *Any simply-compounded forward rate spanning a time interval ending in T_i is a martingale under the $\mathbb{T}_M$-forward measure, i.e.,*

$$\left(\frac{P(t,T_i)}{P(t,T_M)}\right)F(t;T_{i-1},T_i) = E^{\mathbb{T}_M}\left[\left(\frac{P(u,T_i)}{P(u,T_M)}\right)F(u;T_{i-1},T_i)|\mathcal{F}_t\right], \qquad (29)$$

for each $0 \leq t \leq u \leq T_{i-1} < T_i \leq T_M$. In particular, if $u = T_{i-1}$ we have that $F(T_{i-1};T_{i-1},T_i) = L(T_{i-1},T_i)$ is the forward rate spanning the interval $[S,T_1]$ is the $Q^{\mathbb{T}_M}-$expectation for the future simply compounded spot rate at time T_{i-1} for the maturity T_i, i.e.,

$$\frac{P(t,T_i)}{P(t,T_M)}F(t;T_{i-1},T_i) = E^{\mathbb{T}_M}\left[L(T_{i-1},T_i)\frac{P(T_{i-1},T_i)}{P(T_{i-1},T_M)}|\mathcal{F}_t\right], \qquad (30)$$

for each $0 \leq t \leq T_{i-1} < T_i \leq T_M$.

Proof. The above expression may be written as

$$\left(\frac{P(t,T_i)}{P(t,T_M)}\right)F(t;T_{i-1},T_i) = (1 + \tau(T_{i-1},T_M)F(t;T_{i-1},T_M)) - \frac{(1 + \tau(T_i,T_M)F(t;T_i,T_M))}{\tau(T_{i-1},T_i)}. \qquad (31)$$

As $F(t;T_{i-1},T_M)$ and $F(t;T_i,T_M)$ are martingales, so it is the above expression. ∎

There is a different proof in the Appendix A. Eq. (29) can be written in a different way as,

$$\frac{P(t,T_{i-1}) - P(t,T_i)}{P(t,T_M)} = E^{\mathbb{T}_M}\left[\frac{P(u,T_{i-1}) - P(u,T_i)}{P(u,T_M)}\Big|\mathcal{F}_t\right]. \qquad (32)$$

3.5. The method for the i-th caplet

Remember that

$$p_t^{\text{caplet}_i} = P(t,T_M)E^{\mathbb{T}_M}\left[\frac{Y\tau_i\left(L(T_{i-1},T_i) - K\right)^+}{P(T_i,T_M)}\Big|\mathcal{F}_t\right]. \qquad (33)$$

In Eq. (33), we need to find the distribution of $P(T_i,T_M)$ and $P(T_{i-1},T_i)$. The restriction that those random variables should satisfy is given by Proposition 1

$$\frac{P(t,T_i)}{P(t,T_M)}F(t;T_{i-1},T_i) = E^{\mathbb{T}_M}\left[L(T_{i-1},T_i)\frac{P(T_{i-1},T_i)}{P(T_{i-1},T_M)}|\mathcal{F}_t\right]. \qquad (34)$$

In order to consider the above restriction we need also to find the distribution of $P(T_{i-1},T_M)$. Summing up, we need the joint distributions of $P(T_{i-1},T_i)$, $P(T_{i-1},T_M)$, and $P(T_i,T_M)$.

The method can be defined now. Suppose we know the distribution of those three variables and, moreover, assume that it is approximated by a discrete probability function with N states denoted by π_k. Find the distribution $\overline{\pi}_k$ in the $\mathbb{T}_M-$forward measure that solves the following minimization problem

$$\overline{\pi}_k = \arg \min_{\overline{\pi}_k} I\left(\overline{\pi}_k, \pi_t\right) \tag{35}$$

$$s.t.$$

$$\sum_{k=1}^{N} \overline{\pi}_k = 1, \tag{36}$$

$$\overline{\pi}_k > 0 \text{ for } k = 1, 2, \ldots, N, \tag{37}$$

$$\sum_{k=1}^{N} \overline{\pi}_k \left(\left(\frac{P_k(T_{i-1}, T_i)}{P_k(T_{i-1}, T_M)}\right) L_k(T_{i-1}, T_i)\right) = \left(\frac{P(t, T_i)}{P(t, T_M)}\right) F\left(t; T_{i-1}, T_i\right), \tag{38}$$

$$\sum_{k=1}^{N} \overline{\pi}_k \left(L_k(T_{i-1}, T_M)\right) = F\left(t; T_{i-1}, T_M\right), \tag{39}$$

$$\sum_{k=1}^{N} \overline{\pi}_k \left(L_k(T_i, T_M)\right) = F\left(t; T_i, T_M\right). \tag{40}$$

3.6. The method for caps: all caplets together

The optimization problem becomes

$$\overline{\pi} = \arg \min_{\overline{\pi}_k} I\left(\overline{\pi}, \pi\right) \tag{41}$$

$$s.t.$$

$$\sum_{k=1}^{N} \overline{\pi}_k = 1 \tag{42}$$

$$\overline{\pi}_k > 0 \qquad\qquad\qquad \text{for } k = 1, 2, \ldots, N, \tag{43}$$

$$\sum_{k=1}^{N} \overline{\pi}_k \left(\left(\frac{P_k(T_{i-1}, T_i)}{P_k(T_{i-1}, T_M)}\right) L_k(T_{i-1}, T_i)\right) = \left(\frac{P(t, T_i)}{P(t, T_M)}\right) F\left(t; T_{i-1}, T_i\right) \quad \text{for } i = 1, \ldots, M-1, \tag{44}$$

$$\sum_{k=1}^{N} \overline{\pi}_k \left(L_k(T_i, T_M)\right) = F\left(t; T_i, T_M\right) \qquad\qquad \text{for } i = 01, \ldots, M-1. \tag{45}$$

Note that the number of equality restrictions related to the proposition 1 (i.e., not including the equality $\sum_{k=1}^{N} \overline{\pi}_k = 1$) is two times the number of caplets minus 1.

3.7. Cressie–Read and duality solution

The previous section used the KLIC as the divergence criteria, but others criteria may be used as well. Note that the above problem is N dimensional with N very high if we want (π_k) to be a good approximation; then, it might be computationally very expensive. Nonetheless, the dual problem is much simpler.

Any divergence criteria could be used in principle, but we need to express it in a convenient way in order to be feasible. Almeida and Garcia (2012), building on Borwein and Lewis (1991), show that it is feasible to use Cressie–Read family function as a divergence criteria in some econometric problems. Such a result is used

in Almeida et al. (2019) to generalize[11] Stutzer (1996). In the same spirit, we will use here the Cressie–Read family showing how to use it in the dual problem.

The Cressie–Read divergence criteria can be written as

$$CR_\gamma(\overline{\pi}, \pi) = \sum_{k=1}^{N} \pi_k \frac{\left(\frac{\overline{\pi}_k}{\pi_k}\right)^{\gamma+1} - 1}{\gamma(\gamma+1)}. \tag{46}$$

Using this function as the objective function, we find the dual problem as (see Appendix Appendix B and Borwein and Lewis (1991))

$$\left(\widehat{\lambda}, \widehat{\mu}\right) = \arg \sup_{(\lambda, \mu) \in \Lambda} -\frac{1}{\gamma+1} \sum_{k=1}^{N} \pi_k \left(1 + \sum_i \lambda_i A_{i,k} + \sum_i \mu_i B_{i,k}\right)^{\frac{\gamma+1}{\gamma}}, \tag{47}$$

where

$$A_{i,k} = \left(\frac{1 - P_k(i-1, i)}{P_k(i-1, M)}\right) - \frac{P(t, T_{i-1}) - P(t, T_i)}{P(t, T_M)}, \tag{48}$$

$$B_{i,k} = \left(\frac{1}{P_k(i, M)} - 1\right) - \frac{P(t, T_i) - P(t, T_M)}{P(t, T_M)}, \tag{49}$$

$$\Lambda = \left\{(\lambda, \mu)| \left(1 + \sum_i \lambda_i A_{i,k} + \sum_i \mu_i B_{i,k}\right) > 0 \text{ for all } k\right\}. \tag{50}$$

Moreover, we have for the $\mathbb{T}_M-$forward measure via the following formula

$$\overline{\pi}_k = \pi_k \frac{\left(1 + \sum_i \lambda_i A_{i,k} + \sum_i \mu_i B_{i,k}\right)^{1/\gamma}}{\sum_{k=1}^{N} \pi_k \left(1 + \sum_i \lambda_i A_{i,k} + \sum_i \mu_i B_{i,k}\right)^{1/\gamma}}. \tag{51}$$

Note that we diminish the dimensionality of the problem from N to two times the number of caplets minus 1. The KLIC is obtained as a particular case of the Cressie–Read family when $\gamma \to 0$ and the formulas above are a little different. Note that for the last caplet there is only one equality restriction related to bond prices, and the dual optimization problem becomes unidimensional.

3.8. Duality interpretation and the portfolio of forward rate agreements (FRA)

We will interpret the dual problem as a portfolio problem when the investor only has the chance to invest in a *forward rates agreements* (FRA) in t. To see this, we will see that $B_{i,k}$ is the payoff of a FRA and $A_{i,k}$ is the deferred payoff of a FRA at a time T_M. The investor has to decide all the quantities at t, although the deferred payoff related $A_{i,k}$ can be interpreted as a dynamic strategy.[12]

3.9. Quantities $A_{i,k}$ and $B_{i,k}$ as payoffs

Remember that the FRA is a contract that pays the difference between a variable rate and a fixed rate on a notional value. Considering the times t, T_i and T_M, its payoff can be written as

$$Y \tau_{i,M} \left(L_k(T_i, T_M) - K\right). \tag{52}$$

---

[11]Haley and Walker (2009) defines a similar generalization, but they use only three particular cases of the Cressie–Read family function that has an almost closed form solution.

[12]Being more precise, it may be interpreted as a dynamical strategy but the actions are taken independently of any information available after t.

If we make $Y = 1$ and choose $K = F(t; T_i, T_M)$ we have that the FRA has zero price and its payoff is

$$\tau_{i,M} \left(L_k(T_i, T_M) - F(t; T_i, T_M) \right). \tag{53}$$

Note that

$$B_{i,k} = \left(\frac{1}{P_k(i, M)} - 1 \right) - \frac{P(t, T_i) - P(t, T_M)}{P(t, T_M)} \tag{54}$$

$$B_{i,k} = \tau_{i,M} \left(L_k(T_i, T_M) - F(t; T_i, T_M) \right), \tag{55}$$

i.e., $B_{i,k}$ is the payoff of a FRA that pays at T_M whose price in t is zero, as we claimed in the first paragraph of this subsection.

For the deferred payoff, consider the dates t, T_{i-1} and T_i with the payoff deferred from T_i to T_M. For the fixed payment, we can defer it in a simpler way

$$\tau_{i-1,i} F(t; T_{i-1}, T_i) \frac{P(t, T_i)}{P(t, T_M)} = \tau_{i-1,i} F(t; T_{i-1}, T_i) \left(1 + \tau_{i,M} F(t; T_i, T_M) \right), \tag{56}$$

i.e., the FRA's fixed leg payoff is deferred to T_M. On the other hand, if one wants to transfer the variable leg it is more complicated. The payoff received in T_i but known in T_{i-1} can be transferred to T_M without risk with the forward rate $F_k(T_{i-1}; T_i, T_M)$. It is necessary to use the rate only known in T_{i-1} because the payoff is not known before. The variable leg payoff can be written as $\tau_{i-1,i} L_k(T_{i-1}, T_i)$ and the deferred payoff can be written as

$$\tau_{i-1,i} L_k(T_{i-1}, T_i) \left(\frac{P_k(T_{i-1}, T_i)}{P_k(T_{i-1}, T_M)} \right) = \tau_{i-1,i} L_k(T_{i-1}, T_i) \left(1 + \tau_{i,M} F_k(T_{i-1}; T_i, T_M) \right). \tag{57}$$

In this way the FRA contract whose date are t, T_{i-1} and T_i has the payoff

$$\tau_{i,M} \left(L_k(T_{i-1}, T_i) - F(t; T_{i-1}, T_i) \right), \tag{58}$$

and a defferred payoff

$$\tau_{i-1,i} L_k(T_{i-1}, T_i) \left(1 + \tau_{i,M} F_k(T_{i-1}; T_i, T_M) \right) - \tau_{i-1,i} F(t; T_{i-1}, T_i) \left(1 + \tau_{i,M} F(t; T_i, T_M) \right), \tag{59}$$

that is equal to

$$\tau_{i-1,i} \left[L_k(T_{i-1}, T_i) \left(\frac{P_k(T_{i-1}, T_i)}{P_k(T_{i-1}, T_M)} \right) - F(t; T_{i-1}, T_i) \frac{P(t, T_i)}{P(t, T_M)} \right] = \tau_{i-1,i} A_{i,k}. \tag{60}$$

3.10. The portfolio problem for CRRA investor

Note that if we make

$$\lambda_i = \frac{\overline{\lambda}_i}{W(0) R_f}, \tag{61}$$

$$\mu_i = \frac{\overline{\mu}_i}{W(0) R_f},$$

where $W(0)$ is the initial wealth (i.e. wealth at t) and $R_f = 1/P(t, T_M)$ is the risk-free rate between t and T_M, we have

$$
\begin{aligned}
1 + \sum_i \lambda_i A_{i,k} + \sum_i \mu_i B_{i,k} &= 1 + \sum_i \frac{\overline{\lambda}_i}{W(0)R_f} A_{i,k} + \sum_i \frac{\overline{\mu}_i}{W(0)R_f} B_{i,k} \\
&= \frac{1}{W(0)R_f} \left(W(0)R_f + \sum_i \overline{\lambda}_i A_{i,k} + \sum_i \overline{\mu}_i B_{i,k} \right),
\end{aligned}
\tag{62}
$$

and identifying the final wealth (i.e. wealth at T_M) as

$$
W(1) = W(0)R_f + \sum_i \overline{\lambda}_i A_{i,k} + \sum_i \overline{\mu}_i B_{i,k},
\tag{63}
$$

we can interpret $\overline{\lambda}_i$ and $\overline{\mu}_i$ as the positions in FRAs whose payoff is $A_{i,k}$ and in FRAs whose deferred payoff is $B_{i,k}$. Note that this portfolio has zero value at t because the FRAs mentioned have price zero at t.

Finally, if we consider the portfolio problem of a CRRA investor

$$
\max_{\overline{\lambda}, \overline{\mu}} E\left[u(W(1))\right],
\tag{64}
$$

subject to

$$
\begin{aligned}
W_k(1) &= W(0)R_f + \sum_i \overline{\lambda}_i A_{i,k} + \sum_i \overline{\mu}_i B_{i,k}, \\
W_k(1) &> 0 \text{ for all states},
\end{aligned}
$$

where

$$
u(W(1)) = -\frac{1}{\gamma + 1} \left(W(1) \right)^{\frac{\gamma+1}{\gamma}}.
$$

Note that the only difference in the above objective function $E\left[u(W(1))\right]$ and the dual objective function in the previous subsection is a positive constant $(1/W(0)R_f)^{\frac{\gamma+1}{\gamma}}$. Moreover, the positive constraint in wealth coincides with the set Λ in the dual problem.

4. Numerical application: exact derivatives pricing under the Heath et al. (1992) model

The previous section provided the Radon–Nikodym (RD) derivative for the forward measure with respect to empirical measure; some pricing models provide this derivative as well. For instance, in continuous time models, the Girsanov Theorem provides such a derivative. The goal of the present section is to compare the derivatives implied by such models and by the method.

4.1. HJM model, forward rates and no-arbitrage conditions

Following HJM, we define the instantaneous forward rate and consider its stochastic differential equation as follows

$$
\begin{aligned}
f(t,T) &= -\frac{\partial}{\partial T} P(t,T), \tag{65} \\
df(t,T) &= \alpha(t,T)\,dt + \sigma(t,T)\,dW_t, \tag{66}
\end{aligned}
$$

13

where W_t is the Wiener process. This implies that $P(t,T) = \exp\left\{-\int_t^T f(t,u)du\right\}$ and that the bond prices evolves as

$$dP(t,T) = P(t,T)\left[R(t) - \alpha^*(t,T) + \frac{1}{2}(\sigma^*(t,T))^2\right] - \sigma^*(t,T)P(t,T)dW_t, \tag{67}$$

where

$$\alpha^*(t,T) = \int_t^T \alpha(t,u)\,du, \quad \sigma^*(t,T) = \int_t^T \sigma(t,u)\,du. \tag{68}$$

HJM showed a sufficient conditions for a model to have no arbitrage is the existence of a process $\Theta(t)$ such that

$$\alpha(t,T) = \sigma(t,T)\left[\sigma^*(t,T) + \Theta(t)\right], \tag{69}$$

and that the Wiener processes in the risk-neutral measure and in the $\mathbb{T}_\beta-$forward measure are, respectively,

$$\widetilde{W}_t = \int \Theta(t)\,dt + dW_t, \tag{70}$$

$$W_t^{T_\beta} = \int \left[\Theta(t) + \sigma^*(t,T_\beta)\right]dt + dW_t. \tag{71}$$

It is convenient to our purpose to consider the SDE for the forward measure as[13]

$$dF(t;T,T_\beta) = \widetilde{\gamma}(t,T)\,F(t;T,T_\beta)\,dW_t^{T_\beta}. \tag{72}$$

It implies that the SDE for empirical measure is

$$dF(t;T,T_\beta) = \widetilde{\gamma}(t,T)\left[\Theta(t) + \sigma^*(t,T_\beta)\right]F(t;T,T_\beta)\,dt + \widetilde{\gamma}(t,T)\,F(t;T,T_\beta)\,dW_t \tag{73}$$

and that $\widetilde{\gamma}(t,T)$ should be equal to

$$\gamma(t,T) = \frac{1 + \tau_\beta F(t;T,T_\beta)}{\tau_\beta F(t;T,T_\beta)}\left[\sigma^*(t,T_\beta) - \sigma^*(t,T)\right]. \tag{74}$$

Finally, the RD given by Girsanov Theorem is

$$Z_t^{T+\delta} = \exp\left\{-\int_0^t \left[\Theta(u) + \sigma^*(u,T_\beta)\right]dW_u - \frac{1}{2}\int_0^t \left(\Theta(u) + \sigma^*(u,T_\beta)\right)^2 du\right\}, \tag{75}$$

and it is convenient to rewrite here the RD provided by the method

$$Z^{\gamma,lastCaplet} = \frac{\left(1 + \mu_i\left[L(T,T+\delta) - F(t,T,T+\delta)\right]\right)^{1/\gamma}}{E\left[\left(1 + \mu_i\left[L(T,T+\delta) - F(t,T,T+\delta)\right]\right)^{1/\gamma}\right]}. \tag{76}$$

4.2. An exact result for the last caplet

We have the following result

---

[13]The existence of $\gamma(t,T)$ is granted by the Martingale Existence Theorem and the positivity of $F(t;T,T_\beta)$.

Lemma 2. *In a model driven by one Wiener Process, if there is only the last caplet and if*

$$\widetilde{\gamma}(t, T_{\beta-1}) = \frac{\alpha(t, T_\beta)}{\sigma(t, T_\beta)}, \tag{77}$$

then the Radon–Nikodym derivatives for the continuous model are the same as the ones provided by the method when using the element in the Cressie–Read family function defined[14] by $\gamma = -1$ in Eq. (46).

4.3. Forward LIBOR model

The forward rate evolves in the empirical measure for the LIBOR market model as

$$dF(t, T, T_\beta) = \widetilde{\gamma}(t, T)\left[\Theta(t) + \sigma^*(t, T_\beta)\right]F(t; T, T_\beta)\,dt + \widetilde{\gamma}(t, T)F(t; T, T_\beta)\,dW_t. \tag{78}$$

Considering only the last caplet, $\widetilde{\gamma}(t, T_{\beta-1})$ as constant and $\widetilde{\gamma}(t, T_{\beta-1}) = \frac{\alpha(t, T_\beta)}{\sigma(t, T_\beta)}$, we have that $F(t, T, T_\beta)$ follows a lognormal distribution and we have a case similar to the Black–Scholes model and the Black caplet formula follows. As the condition in the lemma is satisfied, the method provides the same RD derivative as the model and the price will be the same.

5. Empirical application: implicit entropic risk measures

In this section we use $\hat{\pi}, \lambda, \mu$ entropic solutions to the main dual optimization problem (Eq. (47)) in Section 3, jointly with the cap and caplet prices in Eq. (28) and Eq. (33), to extract the *entropic risk content* of the US interest rate derivatives market: the implicit *entropic price call-premium* of *caps* are as in Eq. (9).

5.1. Data and zero-coupon interest rate term structure estimation

Zero coupon rates are constructed by using daily close prices of: (i) the US interest rates of the LIBOR curve, with maturities that range from 1-day (overnight) to 6 months for the short-term part of the curve, and (ii) the US interest rate swap curve for maturities that range from 1 year to 40 years, for the medium- to the long-term part of the term structure. As the fixed-income markets are in general over-the-counter (OTC), we use the data collected and provided by the Bloomberg platform. A detailed description of the instruments is in Table 1. The data spans from May 10, 2005 to August 01, 2013 (Figure 1). Bloomberg swap rates are calculated from the Treasury bonds' mid-prices and the quoted swap spreads.

Discount factors and forward rates are estimated from the zero-coupon rates (swap rates). With the discount factors and the forward rates, we estimate the interest rate term structure using a Nelson and Siegel (1987) curve fitting,[15] that will be useful for extrapolating the rates of intermediate maturities. The resulting Nelson and Siegel (1987) fitted interest rate term structure can be used to extrapolate the inputs of the main dual optimization problem (Eq. (47)): the forward price terms $P_k(i-1, i), P_k(i, M)$ and price terms $P(t, T_{i-1}), P(t, T_i), P(t, T_M)$; with $i \in \{1, \ldots, M\}$ the caplets maturities considered in the valuation, M the cap maturity, T_i the time of the forward measures from the i-th maturity term, k the k-th scenario, and $P(\cdot)$ bond prices.

---

[14]Remember that γ is a constant that defines the element in the Cressie–Read family function in Eq. (46) and $\widetilde{\gamma}(t, T_{\beta-1})$ is a process related to the SDE for the forward rate.

[15]The Nelson and Siegel (1987) yield curve modeling provides a smooth interpolation method of interest rate term structure that has no-arbitrage conditions. Although it is not recent, it continues to be one of the most used methods in industry and academia for modeling the interest rate term structure, such as in Diebold and Li (2006) and Diebold et al. (2008).

5.2. Implied risk-neutral volatility estimation

The second dataset used in this research is the market risk-neutral interest rate implied volatility, extracted from the caps/floors volatility using a method named the *interest rate volatility cube* from Hagan and Konikov (2004). This risk-neutral market implied volatility is used within the Black (1976) framework to calculate the market implicit call price, C_m in Eq. (9). The market *interest rate volatility cube* volatility is extracted from the Bloomberg terminal, for the caps/floors with strikes of 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 11%, 12%, 13%, and 14%, with maturities that range from 1 to 10 years. Figure 2 shows the interest rate market implied volatility from the interest rate volatility cube for June 21, 2005.

5.3. Probability scenarios: the interest rate term structure probability grid

For the estimation of the implied entropic risk-neutral estimation, we need to construct a probability grid of the interest rate term structure. Every path of the grid represents one realization $\bar{\pi}_k, k \in \{1, \dots, N\}$. Stutzer (1996) and Stutzer and Chowdhury (1999) used historical realizations of stock prices to construct the scenarios. We follow this procedure by generating a grid of scenarios with the last year ($N = 252$ business days) interest rate term structure observations (see Figure 3a); but in a interest rate environment, the dynamic of the term structure provides more information than the dynamic of the stock prices; for this reason, and to gain modeling power, we generate other two grid scenarios for the *canonical valuation* of options: (i) one plain grid with non-arbitrage restrictions for the physical measure (see Figure 3b), and (ii) one arbitrage-free grid using Nelson and Siegel (1987) and Svensson (1994) restrictions (see Figure 3c).[16]

Table 1: **Interest Rate Market Data**

This table displays the different fixed-income instruments used for calculating the interest rate term structure curve used in the valuation of interest rate swaps (IRS).

Interest rate description	Tenor	Bloomberg Code	Mean	Std. Dev.	Skewness	Kurtosis
US dollar *LIBOR*	1 day	US00O/N Index	2.0536	2.1399	0.445	1.4368
	1 week	US0001W Index	2.1152	2.1405	0.417	1.4139
	1 month	US0001M Index	2.1687	2.1437	0.4062	1.4091
	2 month	US0002M Index	2.2546	2.1255	0.3821	1.3998
	3 month	US0003M Index	2.3242	2.1079	0.3666	1.3957
	6 month	US0006M Index	2.4899	2.0184	0.336	1.3918
US dollar swap rate	1 year	USSW1 Curncy	2.3865	2.064	0.3777	1.3967
	2 years	USSW2 Curncy	2.5145	1.9397	0.3169	1.3983
	3 years	USSW3 Curncy	2.7094	1.8394	0.2058	1.416
	4 years	USSW4 Curncy	2.9139	1.7408	0.1055	1.4551
	5 years	USSW5 Curncy	3.1075	1.6431	0.0273	1.5006
	6 years	USSW6 Curncy	3.2805	1.5523	-0.0333	1.542
	7 years	USSW7 Curncy	3.4288	1.4733	-0.0798	1.575
	8 years	USSW8 Curncy	3.5514	1.4084	-0.1129	1.5989
	9 years	USSW9 Curncy	3.6546	1.3551	-0.138	1.6164
	10 years	USSW10 Curncy	3.7441	1.3095	-0.1569	1.63
	12 years	USSW12 Curncy	3.8904	1.2367	-0.1879	1.6494
	15 years	USSW15 Curncy	4.0375	1.1697	-0.2195	1.6677
	20 years	USSW20 Curncy	4.1506	1.1211	-0.2347	1.6698
	25 years	USSW25 Curncy	4.1995	1.0951	-0.2424	1.6705
	30 years	USSW30 Curncy	4.2255	1.0766	-0.2485	1.6762
	35 years	USSW35 Curncy	4.2324	1.0707	-0.2586	1.6835
	40 years	USSW40 Curncy	4.2326	1.0657	-0.2567	1.6792

---

[16]Gerstner and Griebel (2003) and Reisinger (2013) are examples of multidimensional grid approaches with complexity reduction for integral valuations in multidimensional environments.

These grids provide only the initial setup; every line path is considered a scenario, where all scenarios have the same probability: $1/N$. The solution of the dual optimization problem (Eq. (47)) generates a entropic version of $\bar{\pi}_k$ for every grid; the grids only limit the possible path scenarios, but they do not limit the probabilities assignment. Observing Figure 3 we note that the plain grid has the greater flexibility and maximum-entropy,[17] it has the initial problem of having paths with "arbitrage" possibilities. Given that market interventions, such as the *Quantitative Easing* provide means for statistical arbitrage, we include this grid. In the case of the *plain grid* and the Svensson (1994) *grid*, $N = 100,000$. To calculate the *grid* risk-neutral densities, we used MATLAB® on an Intel® Xeon® X5670-based cluster with 148 threads.
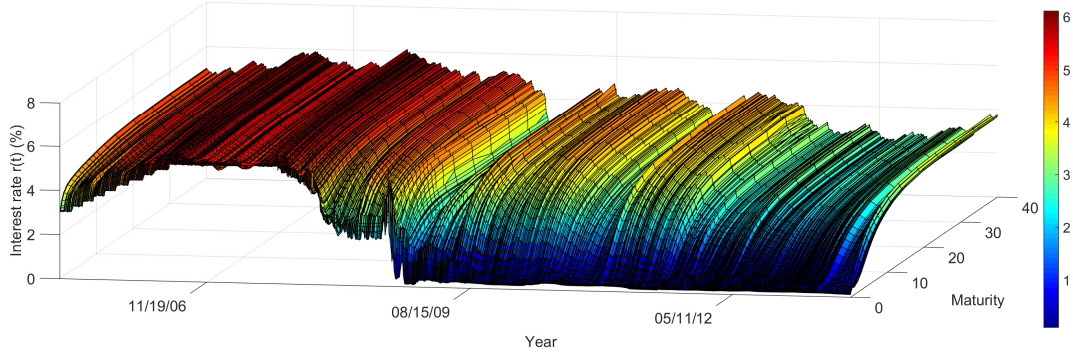
(a) US *LIBOR* and swap rates surface evolution for different maturities.
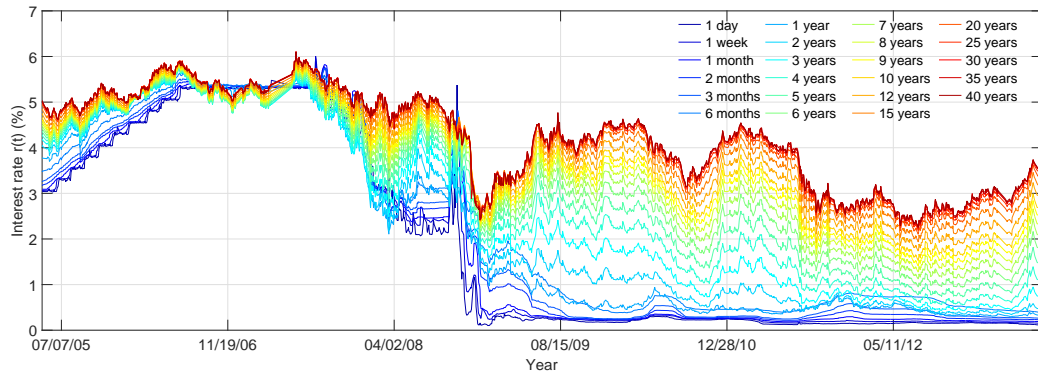
(b) US *LIBOR* and swap rates evolution in time.

Figure 1: This figure shows the *LIBOR* and swap rates evolution from May 2005 to August 2013. The historical data of *LIBOR* rates go from spot to 6-month maturities. The swap rates are from 1 year to 40-year maturities. The rates are extracted from Bloomberg.
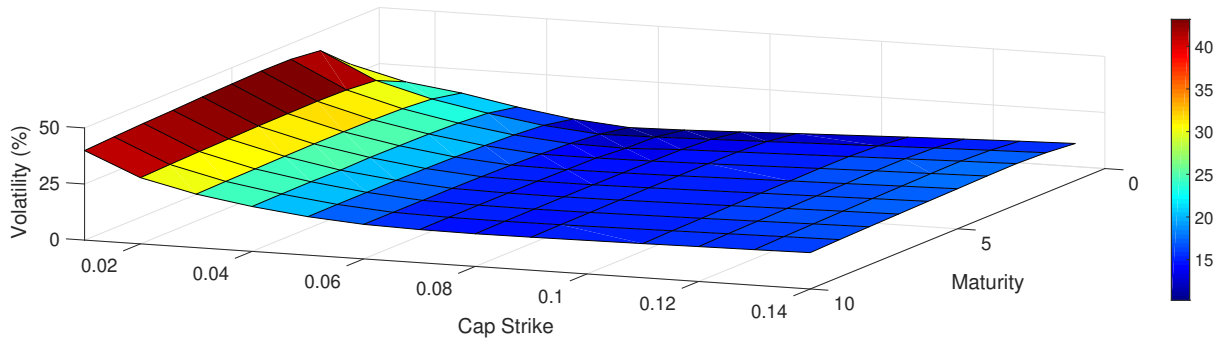

Figure 2: Volatility surface of caps/floors for the period of June 21, 2006.

---

[17]See Gzyl and Mayoral (2008) and Gzyl and Mayoral (2012) for uses of the maximum entropy in physical risk measurement

5.4. Results

Figure 4 shows some results for the estimated *entropic risk-neutral density* by solving the optimization problem (Eq. (47)). We use $\gamma = 1$ that represents the classical Euclidean likelihood (Euclidean divergence). The entropic risk-neutral probabilities were interpolated with a polynomial spline to plot smooth surfaces (probabilities obtained are discrete, in a grid of $N = 252$ scenarios for the historical grid cases, Figure 4a, and $N = 100,000$ for the *plain* and Svensson (1994) *grid* case, Figures 4b and 4c). We include a fourth estimated implicit risk-neutral density (see Figure 4d), the Li and Zhao (2009) nonparametric risk-neutral density with polynomial distances. Densities in Figure 4 are from June 21, 2006. We observe that the initial *grid* setup determines the final outcome: initial restrictions mold solutions of the optimization problem (Eq. (47)).
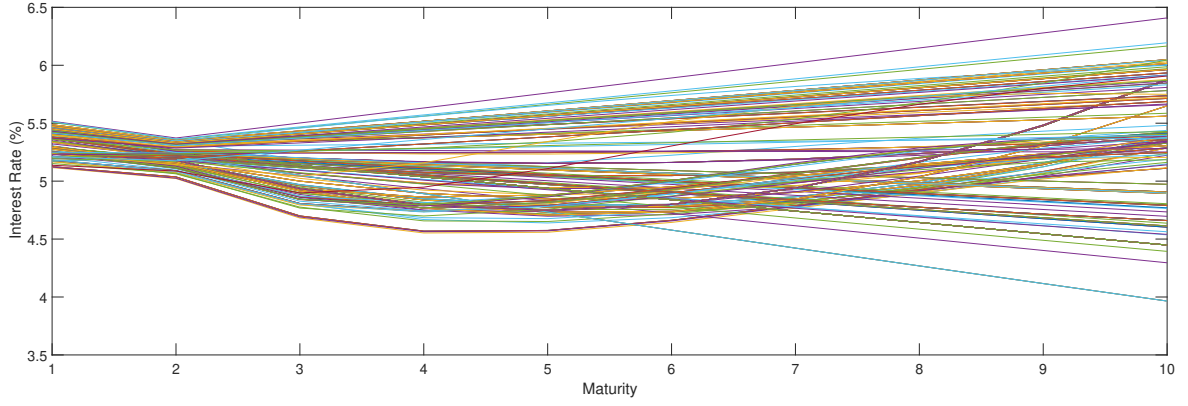
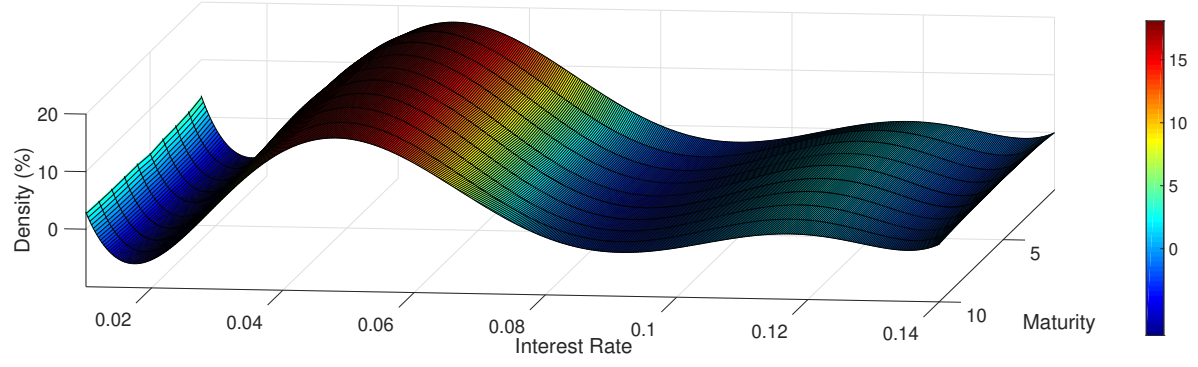(a) Probability historical data grid ($N = 252$ days).
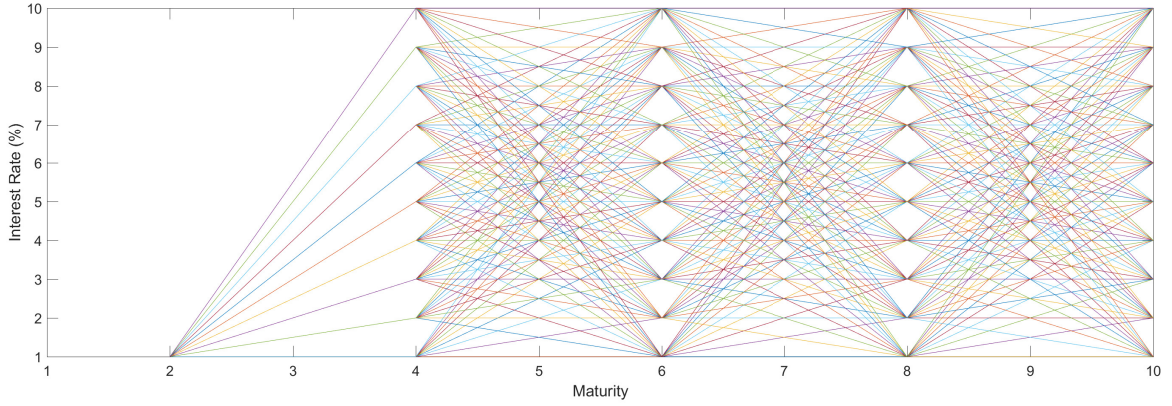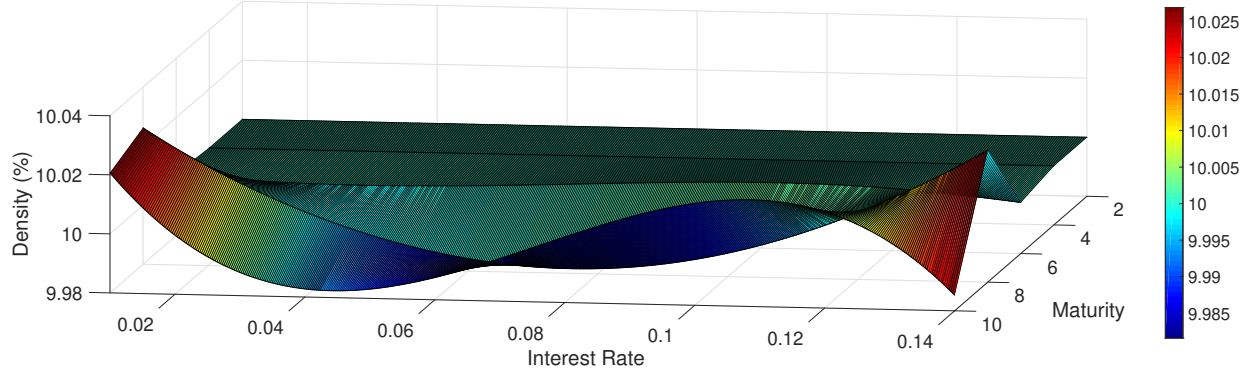

(b) Probability flat grid.
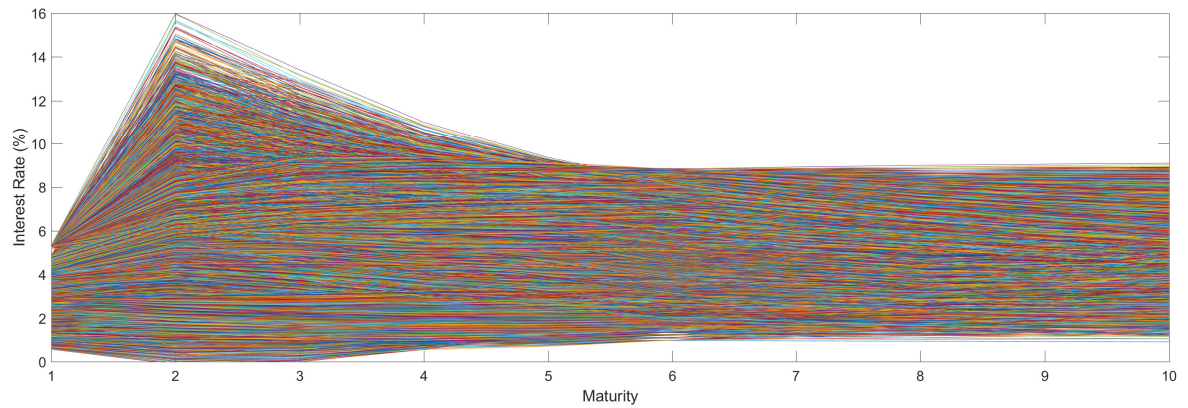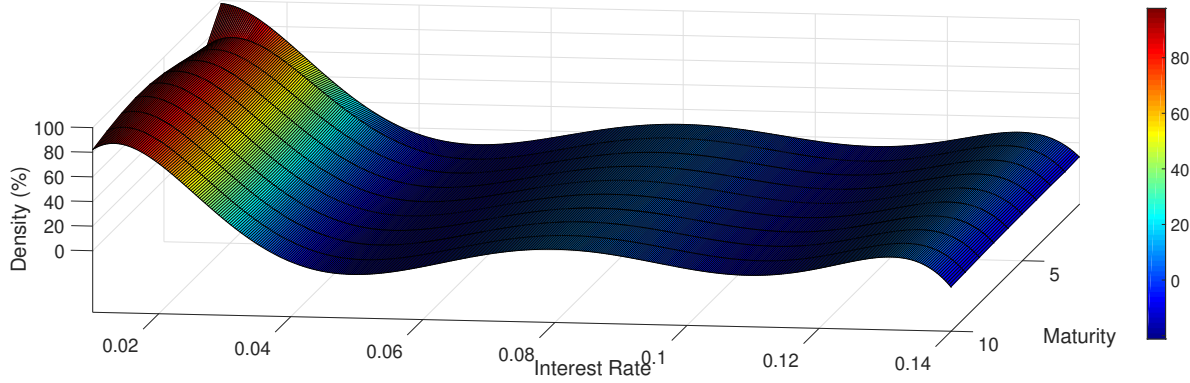

(c) Probability Svensson grid.
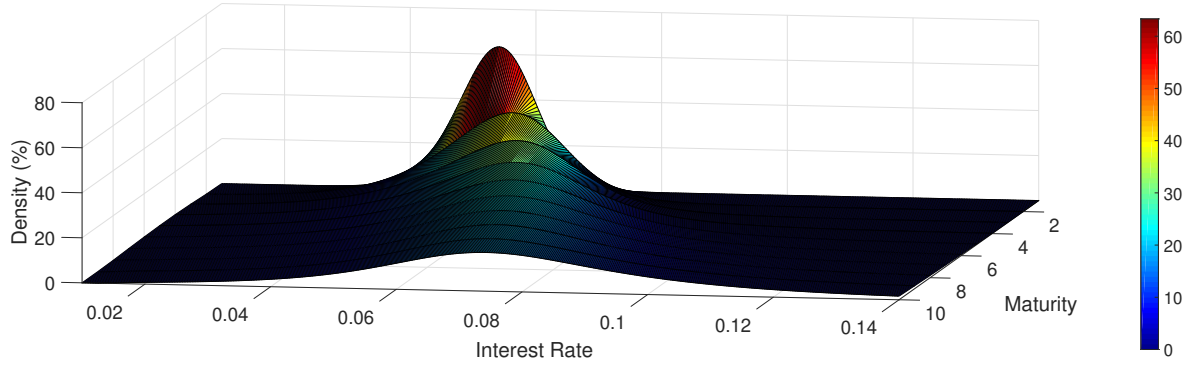
Figure 3: Probability scenario grids.

18

(a) Implied historical *grid* risk-neutral density.


(b) Implied plain *grid* risk-neutral density.

(c) Implied Svensson's (1994) grid risk-neutral density.

(d) Implied Li and Zhao's (2009) risk-neutral density.

Figure 4: Implied risk-neutral densities for June 21, 2006, resulting from Eq. (51) for the different *grid* methods. The Li and Zhao's (2009) risk-neutral density is estimated by using exponential polynomials restrictions. Pdf's are marginal densities for every maturity (but not the joint distribution), and they are standardized to add 1.

The *plain grid* provides the initial setup with greater maximum-entropy, and the final risk-neutral density has the greater maximum-entropy, with minor differences in the long-term maturity caplets (10 years). The *historical grid* solution preserves the initial concentration of the densities between 4% and 6%: density values observations close to zero and to 12% are the result of the polynomial approximation, not the observed values as they are not possible due to the initial path restrictions. The Svensson (1994) *grid* solution concentrates the density on the lower interest rates: Svensson (1994) no-arbitrage constraints consider the long-term forward rates (close to 5%) and limit the possibility of the market having higher interest rates. This result is interesting for policymakers when analyzing no-arbitrage conditions in their monetary policy decisions.
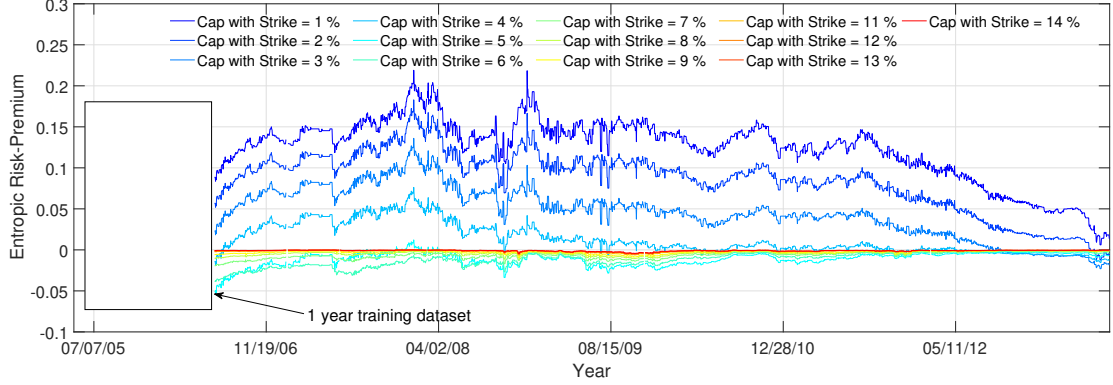
(a) Historical Data Grid (Stutzer, 1996), $\gamma = 1$.
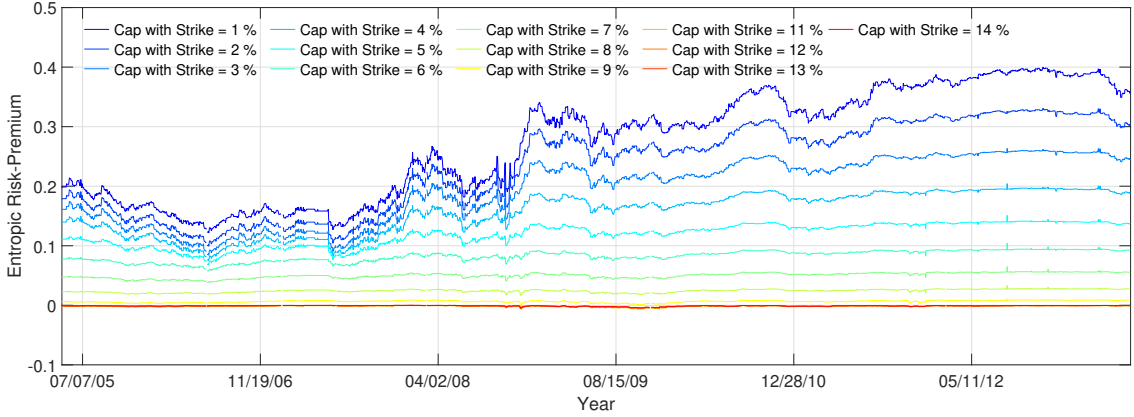
(b) Plain Grid, $\gamma = 1$.
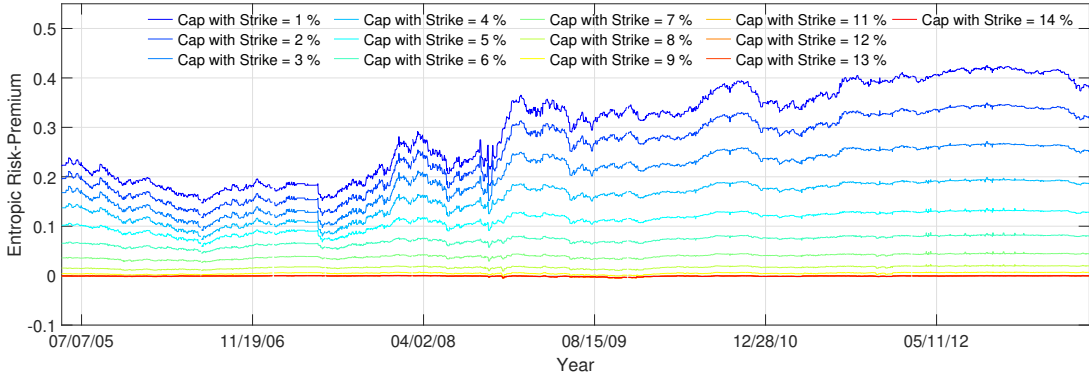
(c) Svensson Grid, $\gamma = 1$.

Figure 5: *Implicit entropic price cap premium.* This figure shows the result of Eq. (9) for the different scenarios/*grid* methods. Data is from the US interest rate derivatives markets from May 10, 2005 to August 01, 2013.

Li and Zhao's (2009) risk-neutral density is closer to a exponential distribution, given the reliance on the polynomial for physical and risk-neutral distance measurement. Figure 4 shows the results for one day. To present the results for the full period (May 2005 to August 20013), we present the *implicit entropic price call-premium*.

Figure 5 shows the *implicit entropic price call-premium* from Eq. (9), using the *historical grid* (Figure 5a), the *plain grid* (Figure 5b), and the Svensson (1994) *grid* (Figure 5c). We observe that when using the *historical grid*, the *entropic call premium* of lower strike caps (between 1% and 4%) peaks during the Bear Stearns and Lehman Brother defaults (March and October 2008), and is reduced towards the end of the period analyzed (August 2013); this finding is consistent with the empirical observation of the interest rate term structure during the crisis period, and it provides a useful method for risk managers that try to assess the *physical* risk measure. Conversely, the *plain* and Svensson (1994) *grids* results show that the *entropic call premium* increases during the *Quantitative Easing* implementation after the 2007/2008 financial crisis; this result is consistent with the arbitrage opportunities that interest rate traders have given the market intervention in the short-term curve for such a long period; this result is useful for monetary policymakers to assess the entropic risk-neutral premium from which the arbitrageurs can profit with their interventions.

6. Conclusions

In this paper we propose a new method to estimate the forward measure of interest rates. With this in mind, it is possible to price interest rates derivatives securities such as caps. Our method is a generalization of the *canonical valuation* introduced by Stutzer (1996) for the case of interest rates. Our method incorporates derivatives prices if available, providing a more accurate estimate of forward measure. The method can be applied in two different ways in order to price caplets. For the i-th caplet, we may price it using a $\mathbb{T}_M-$ forward measure or using a $\mathbb{T}_i-$ forward measure. The latter is computationally faster but the former incorporates more information from available price.

Our derivative pricing method estimates the risk-neutral density using the concept of "entropic risk". Then, we introduce the definition of *entropic risk-neutral density premium*, that is the difference between the interest rate derivative market prices, and the interest rate derivative price extracted from the calibration of the *risk-neutral entropic risk density*. The *implicit entropic risk* concept provides a new measure for interest rate derivative risk managers, which can be used to incorporate concepts of information theory.

The method is applied in a numerical application, to extract the derivatives prices in the Heath et al. (1992) model, and in an empirical application with US interest rates and derivatives data. Results in the numerical case show that under certain assumptions the model can provide exact results, and in the empirical case that the entropic based risk-neutral density is useful to highlight risks previous to the financial crisis, and the statistical arbitrage burden when policymakers decide to implement *Quantitative Easing* for large periods. Future work may provide further characterization on the method's behavior in a different derivatives market.

References

Ahmadi-Javid, A., 2011. Entropic value-at-risk: A new coherent risk measure. Journal of Optimization Theory and Applications 155, 1105–1123. doi:10.1007/s10957-011-9968-2.

Ahmadi-Javid, A., Fallah-Tafti, M., 2019. Portfolio optimization with entropic value-at-risk. European Journal of Operational Research 279, 225–241. doi:10.1016/j.ejor.2019.02.007.

Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. The Journal of Finance 53, 499–547.

Almeida, C., Ardison, K., Azevedo, R., Freire, G., 2019. An anatomy of generalized entropic estimators in option pricing. SSRN Electronic Journal doi:10.2139/ssrn.2535790.

Almeida, C., Garcia, R., 2012. Assessing misspecified asset pricing models with empirical likelihood estimators. Journal of Econometrics 170, 519–537. doi:10.1016/j.jeconom.2012.05.020.

Almeida, C., Graveline, J.J., Joslin, S., 2011. Do interest rate options contain information about excess returns? Journal of Econometrics 164, 35–44. doi:10.1016/j.jeconom.2011.02.007.

Bali, T.G., Murray, S., 2013. Does risk-neutral skewness predict the cross-section of equity option portfolio returns? Journal of Financial and Quantitative Analysis 48, 1145–1171. doi:10.1017/s0022109013000410.

Barletta, A., de Magistris, P.S., Violante, F., 2019. A non-structural investigation of VIX risk neutral density. Journal of Banking & Finance 99, 1–20. doi:10.1016/j.jbankfin.2018.11.012.

Bekiros, S., Nguyen, D.K., Junior, L.S., Uddin, G.S., 2017. Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets. European Journal of Operational Research 256, 945–961. doi:10.1016/j.ejor.2016.06.052.

Ben-Tal, A., 1985. The entropic penalty approach to stochastic programming. Mathematics of Operations Research 10, 263–279. doi:10.1287/moor.10.2.263.

Bikbov, R., Chernov, M., 2013. Monetary policy regimes and the term structure of interest rates. Journal of Econometrics 174, 27–43. doi:10.1016/j.jeconom.2013.01.002.

Black, F., 1976. The pricing of commodity contracts. Journal of Financial Economics 3, 167–179.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Borwein, J.M., Lewis, A.S., 1991. Duality relationships for entropy-like minimization problems. SIAM Journal on Control and Optimization 29, 325–338. doi:10.1137/0329017.

Brandtner, M., Kürsten, W., Rischau, R., 2018. Entropic risk measures and their comparative statics in portfolio selection: Coherence vs. convexity. European Journal of Operational Research 264, 707–716. doi:10.1016/j.ejor.2017.07.007.

Brigo, D., Mercurio, F., 2006. Interest Rate Models — Theory and Practice. Springer Berlin Heidelberg. doi:10.1007/978-3-540-34604-3.

Chernov, M., Ghysels, E., 2000. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. Journal of Financial Economics 56, 407–458. doi:10.1016/s0304-405x(00)00046-5.

Diebold, F.X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337–364. doi:10.1016/j.jeconom.2005.03.005.

Diebold, F.X., Li, C., Yue, V.Z., 2008. Global yield curve dynamics and interactions: A dynamic Nelson–Siegel approach. Journal of Econometrics 146, 351–363. doi:10.1016/j.jeconom.2008.08.017.

Fabozzi, F.J., Tunaru, R., Albota, G., 2009. Estimating risk-neutral density with parametric models in interest rate markets. Quantitative Finance 9, 55–70. doi:10.1080/14697680802272045.

Fleischhacker, A.J., Fok, P.W., 2015. On the relationship between entropy, demand uncertainty, and expected loss. European Journal of Operational Research 245, 623–628. doi:10.1016/j.ejor.2015.03.014.

Föllmer, H., Knispel, T., 2011. Entropic risk measures: Coherence vs. convexity, model ambiguity and robust large deviations. Stochastics and Dynamics 11, 333–351. doi:10.1142/s0219493711003334.

Frittelli, M., 2000. The minimal entropy martingale measure and the valuation problem in incomplete markets. Mathematical Finance 10, 39–52. doi:10.1111/1467-9965.00079.

Gerstner, T., Griebel, M., 2003. Dimension–adaptive tensor–product quadrature. Computing 71, 65–87. doi:10.1007/s00607-003-0015-5.

Gzyl, H., Mayoral, S., 2008. Determination of risk pricing measures from market prices of risk. Insurance: Mathematics and Economics 43, 437–443. doi:10.1016/j.insmatheco.2008.08.008.

Gzyl, H., Mayoral, S., 2012. Determination of the probability distribution measures from market option prices using the method of maximum entropy in the mean. Applied Mathematical Finance 19, 299–312. doi:10.1080/1350486x.2011.621354.

Hagan, P., Konikov, M., 2004. Interest Rate Volatility Cube: Construction And Use. Technical Report. Bloomberg.

Haley, M.R., Walker, T.B., 2009. Alternative tilts for nonparametric option pricing. Journal of Futures Markets , n/a–n/adoi:10.1002/fut.20445.

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, 77–105. doi:10.2307/2951677.

Hoskisson, R.E., Hitt, M.A., Johnson, R.A., Moesel, D.D., 1993. Construct validity of an objective (entropy) categorical measure of diversification strategy. Strategic Management Journal 14, 215–235. doi:10.1002/smj.4250140305.

Kitamura, Y., 2006. Empirical likelihood methods in econometrics: Theory and practice, in: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics and Econometrics. Cambridge University Press, Cambridge, pp. 174–237. doi:10.1017/cbo9780511607547.008.

Leiss, M., Nax, H.H., 2018. Option-implied objective measures of market risk. Journal of Banking & Finance 88, 241–249. doi:10.1016/j.jbankfin.2017.11.017.

Li, H., Zhao, F., 2006. Unspanned stochastic volatility: Evidence from hedging interest rate derivatives. The Journal of Finance 61, 341–378.

Li, H., Zhao, F., 2009. Nonparametric estimation of state-price densities implicit in interest rate cap prices. Review of Financial Studies 22, 4335–4376. doi:10.1093/rfs/hhp025.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curves. The Journal of Business 60, 473–489.

Pichler, A., Schlotter, R., 2019. Entropy based risk measures. European Journal of Operational Research doi:10.1016/j.ejor.2019.01.016.

Reisinger, C., 2013. Asymptotic expansion around principal components and the complexity of dimension adaptive algorithms, in: Garcke, J., Griebel, M. (Eds.), Sparse Grids and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. pp. 263–276.

Rompolis, L.S., 2010. Retrieving risk neutral densities from European option prices based on the principle of maximum entropy. Journal of Empirical Finance 17, 918–937. doi:10.1016/j.jempfin.2010.04.007.

Rompolis, L.S., Tzavalis, E., 2008. Recovering risk neutral densities from option prices: A new approach. Journal of Financial and Quantitative Analysis 43, 1037–1053. doi:10.1017/s0022109000014435.

Rompolis, L.S., Tzavalis, E., 2010. Risk premium effects on implied volatility regressions. Journal of Financial Research 33, 125–151. doi:10.1111/j.1475-6803.2010.01265.x.

Rouge, R., Karoui, N.E., 2000. Pricing via utility maximization and entropy. Mathematical Finance 10, 259–276. doi:10.1111/1467-9965.00093.

Shreve, S., 2004. Stochastic Calculus for Finance II: Continuous-Time Models. Springer Finance Textbooks, Springer, Berlin and Heidelberg.

Shuiabi, E., Thomson, V., Bhuiyan, N., 2005. Entropy as a measure of operational flexibility. European Journal of Operational Research 165, 696–707. doi:10.1016/j.ejor.2004.01.033.

Stutzer, M., 1996. A simple nonparametric approach to derivative security valuation. The Journal of Finance 51, 1633–1652.

Stutzer, M., Chowdhury, M., 1999. A simple non-parametric approach to bond futures option pricing. The Journal of Fixed Income 8, 67–76. doi:10.3905/jfi.1999.319245.

Svensson, L.E., 1994. Estimating and Interpreting Forward Interest Rates: Sweden 1992–1994. Technical Report. doi:10.3386/w4871.

Unal, H., Madan, D., Güntay, L., 2003. Pricing the risk of recovery in default with absolute priority rule violation. Journal of Banking & Finance 27, 1001–1025. doi:10.1016/s0378-4266(02)00255-8.

Zhou, R., Cai, R., Tong, G., 2013. Applications of entropy in finance: A review. Entropy 15, 4909–4931. doi:10.3390/e15114909.

Appendix A. Proof of Proposition 1

For the reader's convenience, we will write two useful propositions as can be found in Brigo and Mercurio (2006). The first one asserts that forward rates are martingales under the T-forward measure and the second one relates to the S-measurability of the T-forward.

Proposition 3. *Any simply-compounded forward rate spannig a time interval ending in T is a martingale under the $\mathbb{T}$-forward measure, i.e.,*

$$E^{\mathbb{T}}\left[F(t;S,T)|\mathcal{F}_u\right] = F(u;S,T), \tag{A.1}$$

for each $0 \leq u \leq t \leq S < T$. In particular, the forward rate spanning the interval $[S,T]$ is the $Q^{\mathbb{T}}-$expectation for the future simply compounded spot rate at time S for the maturity T, i.e.,

$$E^{\mathbb{T}}\left[L(S,T)|\mathcal{F}_t\right] = F(t;S,T), \tag{A.2}$$

for each $0 \leq t \leq S < T$.

Proposition 4. *If H is a $\mathbb{T}$-measurable random variable, we have the identity in the risk-neutral measure*

$$E^{\mathbb{Q}}\left[D(t,S)H|\mathcal{F}_t\right] = E^{\mathbb{Q}}\left[\frac{D(t,T)}{P(S,T)}H|\mathcal{F}_t\right], \tag{A.3}$$

for all $t < S < T$, where $D(t,S) = \int_t^S e^{-r_s ds}$.

Now we are in a position to derive the main pricing result:

Proposition 5 (1). *Any simply-compounded forward rate spanning a time interval ending in $\mathbb{T}_1$ is a martingale under the $\mathbb{T}_2$-forward measure, i.e.,*

$$\frac{P(u,T_1)}{P(u,T_2)}F(u;S,T_1) = E^{\mathbb{T}_2}\left[F(t;S,T_1)\frac{P(t,T_1)}{P(t,T_2)}|\mathcal{F}_u\right],$$

for each $0 \leq u \leq t \leq S < T_1 \leq T_2$. In particular, the forward rate spanning the interval $[S,T_1]$ is the $Q^{\mathbb{T}_2}-$expectation for the future simply compounded spot rate at time S for the maturity T_1, i.e.,

$$\frac{P(u,T_1)}{P(u,T_2)}F(u;S,T_1) = E^{\mathbb{T}_2}\left[L(S,T_1)\frac{P(S,T_1)}{P(S,T_2)}|\mathcal{F}_u\right],$$

for each $0 \leq t \leq S < T$.

Proof. Consider a $\mathbb{T}_1-$measurable random variable H. Proposition 2 implies

$$E^{\mathbb{Q}}\left[D(u,T_1)H|\mathcal{F}_u\right] = E^{\mathbb{Q}}\left[\frac{D(u,T_2)}{P(T_1,T_2)}H|\mathcal{F}_u\right], \tag{A.4}$$

where $u \leq T_1 \leq T_2$.

This implies that using the $\mathbb{T}_1-$forward measure on the left and the $\mathbb{T}_2-$forward measure on the rigth we have

$$
\begin{aligned}
P(u,T_1)E^{\mathbb{T}_1}[H|\mathcal{F}_u] & = P(u,T_2)E^{\mathbb{T}_2}\left[\frac{H}{P(T_1,T_2)}|\mathcal{F}_u\right] \\
E^{\mathbb{T}_1}[H|\mathcal{F}_u] & = \frac{P(u,T_2)}{P(u,T_1)}E^{\mathbb{T}_2}\left[\frac{H}{P(T_1,T_2)}|\mathcal{F}_u\right] \\
E^{\mathbb{T}_1}[F(t;S,T_1)|\mathcal{F}_u] & = \frac{P(u,T_2)}{P(u,T_1)}E^{\mathbb{T}_2}[F(t;S,T_1)[1+L(T_1,T_2)\tau(T_1,T_2)]|\mathcal{F}_u] \\
& = \frac{P(u,T_2)}{P(u,T_1)}E^{\mathbb{T}_2}\left[F(t;S,T_1)\frac{P(t,T_1)}{P(t,T_2)}|\mathcal{F}_u\right] \\
\frac{P(u,T_1)}{P(u,T_2)}F(u;S,T_1) & = E^{\mathbb{T}_2}\left[F(t;S,T_1)\frac{P(t,T_1)}{P(t,T_2)}|\mathcal{F}_u\right].
\end{aligned}
$$

∎

Just for convenience, we do the algebra of the alternative expression here. By definition of $F(u;S,T_1)$ we can do

$$
\frac{P(u,T_1)}{P(u,T_2)}F(u;S,T_1) = \frac{P(u,T_1)}{P(u,T_2)}\frac{\frac{1}{\tau(S,T)}(P(u,S)-P(u,T_1))}{P(u,T_1)} = \frac{1}{\tau(S,T)}\left(\frac{P(u,S)}{P(u,T_2)}-\frac{P(u,T_1)}{P(u,T_2)}\right).
$$

Now, using the proposition we have

$$
\left(\frac{P(u,S)}{P(u,T_2)}-\frac{P(u,T_1)}{P(u,T_2)}\right) = E^{\mathbb{T}_2}\left[\left(\frac{P(t,S)}{P(t,T_2)}-\frac{P(t,T_1)}{P(t,T_2)}\right)|\mathcal{F}_u\right].
$$

Appendix B. Dual Problem

For general problem, we use the results in Borwein and Lewis (1991). Nonetheless, when we assume that we have a finite number of states in the optimization the dual problem is much easier. We do the algebra here for the reader convenience.

The primal problem is

$$
\overline{\pi} = \arg\min_{\overline{\pi}_k} CR_\gamma\left(\overline{\pi},\pi\right) \tag{B.1}
$$

$$
s.t.
$$

$$
\begin{aligned}
\sum_{k=1}^{N}\overline{\pi}_k & = 1 \\
\overline{\pi}_k & > 0 && \text{for } k=1,2,\ldots,N, \\
\sum_{k=1}^{N}\overline{\pi}_k\left(\frac{1-P_k(i-1,i)}{P_k(i-1,M)}\right) & = \frac{P(t,T_{i-1})-P(t,T_i)}{P(t,T_M)}, && \text{for } i=1,\ldots,M-1, \\
\sum_{k=1}^{N}\overline{\pi}_k\left(\frac{1}{P_k(i,M)}-1\right) & = \frac{P(t,T_i)-P(t,T_M)}{P(t,T_M)} && \text{for } i=0,1,\ldots,M-1.
\end{aligned}
$$

Realize that $CR_\gamma\left(\overline{\pi},\pi\right)$ is a convex function.

The Lagragian is

$$L = CR_\gamma\left(\overline{\pi}_k, \pi_k\right) - \sum_{k=1}^{N}\left[\sum_i \widetilde{\lambda}_i A_{i,k}\overline{\pi}_k + \sum_j \widetilde{\mu}_j B_{j,k}\overline{\pi}_k\right] - \nu_0\left(\sum_{k=1}^{N}\overline{\pi}_k - 1\right) + \sum_{k=1}^{N}\nu_k\left(-\overline{\pi}_k\right),$$

where

$$A_{i,k} = \left(\frac{1 - P_k(i-1, i)}{P_k(i-1, M)}\right) - \frac{P(t, T_{i-1}) - P(t, T_i)}{P(t, T_M)}$$

$$B_{i,k} = \left(\frac{1}{P_k(i, M)} - 1\right) - \frac{P(t, T_i) - P(t, T_M)}{P(t, T_M)}$$

and

$$\widetilde{\lambda}_i, \widetilde{\mu}_j, \nu_0 \in \Re \text{ for all } i \in \{1, \dots, M\} \text{ and } j \in \{0, \dots, M-1\},$$

$$\mu_k \geq 0 \text{ for all } k.$$

It will be convenient to define

$$\begin{aligned}
\widetilde{\lambda}_i &= \lambda_i \nu_0, \\
\widetilde{\mu}_j &= \mu_j \nu_0.
\end{aligned}$$

Assume that we have a interior solution. This assumption will be justifiedin the end for $\gamma < 0$. In this case the optimal $\overline{\pi}_k$ will be strictly greater than zero and by the Karush–Kuhn–Tucker Theorem (KKT) we can assume

$$\nu_k = 0 \text{ for } k = 1, 2, \dots, N.$$

This will simplify the algebra, but the same steps are valid without this assumption.

The convex conjugate is

$$\phi\left(\lambda, \mu, \nu_0\right) = \inf_{\overline{\pi}_k}\left\{CR_\gamma\left(\overline{\pi}_k, \pi_k\right) - \sum_{k=1}^{N}\left[\nu_0\sum_i \lambda_i A_{i,k}\overline{\pi}_k + \nu_0\sum_j \mu_j B_{j,k}\overline{\pi}_k\right] - \nu_0\left(\sum_{k=1}^{N}\overline{\pi}_k - 1\right)\right\}.$$

Appendix B.0.1. Solving $\overline{\pi}_k$ given λ, μ and ν_0

The first order conditions are

$$\begin{aligned}
\frac{\partial L}{\partial \overline{\pi}_k} &= (\gamma + 1)\frac{1}{\pi_k}(\pi_k)\frac{\left(\frac{\overline{\pi}_k}{\pi_k}\right)^\gamma}{\gamma(\gamma+1)} - \nu_0\sum_i \lambda_i A_{i,k} - \nu_0\sum_j \mu_j B_{j,k} - \nu_0 = 0, \\
\frac{\partial L}{\partial \overline{\pi}_k} &= \frac{\left(\frac{\overline{\pi}_k}{\pi_k}\right)^\gamma}{\gamma} - \nu_0\left(\sum_i \lambda_i A_{i,k} +_0\sum_j \mu_j B_{j,k} + 1\right) = 0, \\
\frac{1}{\gamma}\left(\frac{\overline{\pi}_k}{\pi_k}\right)^\gamma &= \nu_0\left(\sum_i \lambda_i A_{i,k} +_0\sum_j \mu_j B_{j,k} + 1\right),
\end{aligned} \tag{B.2}$$

$$\overline{\pi}_k \;=\; \pi_k\left[\gamma\nu_0\left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1\right)\right]^{\frac{1}{\gamma}}.$$

Realize that [18]

$$\nu_0\left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1\right) > 0 \quad \text{when } \overline{\pi}_k > 0. \tag{B.3}$$

Appendix B.1. The dual function $\phi\left(\lambda,\mu,\nu_0\right)$

Recalling the dual function

$$\phi\left(\lambda,\mu,\nu_0\right) = \inf_{\overline{\pi}_k}\left\{CR_\gamma\left(\overline{\pi}_k,\pi_k\right) - \sum_{k=1}^{N}\left[\nu_0\sum_i \lambda_i A_{i,k}\overline{\pi}_k + \nu_0\sum_j \mu_j B_{j,k}\overline{\pi}_k\right] - \nu_0\left(\sum_{k=1}^{N}\overline{\pi}_k - 1\right)\right\}. \tag{B.4}$$

Using the first order conditions (and maintaining $\overline{\pi}_k$ as the optimal for notational sake)

$$
\begin{aligned}
\phi\left(\lambda,\mu,\nu_0\right) &= \sum_{k=1}^{N}\pi_k\frac{\left(\frac{\overline{\pi}_k}{\pi_k}\right)^{\gamma+1}-1}{\gamma(\gamma+1)} - \sum_{k=1}^{N}\left[\nu_0\sum_i \lambda_i A_{i,k}\overline{\pi}_k + \nu_0\sum_j \mu_j B_{j,k}\overline{\pi}_k + \nu_0\overline{\pi}_k\right] + \nu_0 \\
&= -\frac{1}{\gamma\left(\gamma+1\right)} + \sum_{k=1}^{N}\pi_k\frac{\overline{\pi}_k}{\pi_k}\left[\frac{\left(\frac{\overline{\pi}_k}{\pi_k}\right)^{\gamma}}{\gamma(\gamma+1)} - \nu_0\left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)\right] + \nu_0,
\end{aligned}
\tag{B.5}
$$

remembering Eq. (B.2)

$$
\begin{aligned}
\phi\left(\lambda,\mu,\nu_0\right) &= -\frac{1}{\gamma\left(\gamma+1\right)} + \nu_0 + \\
&\quad \sum_{k=1}^{N}\overline{\pi}_k\left[\frac{\nu_0\left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1\right)}{\left(\gamma+1\right)} - \left(\frac{\gamma+1}{\gamma+1}\right)\nu_0\left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)\right] \\
&= -\frac{1}{\gamma\left(\gamma+1\right)} + \nu_0 + \sum_{k=1}^{N}\overline{\pi}_k\left[-\gamma\frac{\nu_0\left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1\right)}{\left(\gamma+1\right)}\right] \\
&= -\frac{1}{\gamma\left(\gamma+1\right)} + \nu_0 - \\
&\quad \sum_{k=1}^{N}\pi_k\left[\gamma\nu_0\left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1\right)\right]^{\frac{1}{\gamma}}\left[\gamma\frac{\nu_0\left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1\right)}{\left(\gamma+1\right)}\right] \\
&= -\frac{1}{\gamma\left(\gamma+1\right)} + \nu_0 - \sum_{k=1}^{N}\pi_k\frac{1}{\left(\gamma+1\right)}\left[\gamma\nu_0\left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1\right)\right]^{\frac{1}{\gamma}+1}.
\end{aligned}
\tag{B.6}
$$

This expression above is almost the same as in the main text. The only difference is the constant $-1/\gamma\left(\gamma+1\right)$ and the extra variable ν_0.

---

[18] For $\mu_k \neq 0$, we would have

$$\pi_k^{Q*}\mu_k = \left(\frac{\gamma}{n}\right)^{\frac{1}{\gamma}}\left(\lambda_1\left(1 + \lambda\left(R_k - R^f\right)\right) + \mu_k\right)^{\frac{1}{\gamma}}.$$

Appendix B.2. Optimal $\overline{\pi}_k$ given the optimal λ, μ

Given the optimal λ, μ we can obtain the optimal value for ν_0. Using the first order conditions for ν_0

$$\frac{\partial}{\partial \nu_0} \phi(\lambda, \mu, \nu_0) = 1 - \sum_{k=1}^{N} \left(\frac{1+\gamma}{\gamma}\right) (\nu_0)^{\frac{1}{\gamma}} \pi_k \frac{1}{(\gamma+1)} \left[\gamma \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)\right]^{\frac{1}{\gamma}+1} = 0 \quad (B.7)$$

$$1 = (\gamma)^{\frac{1}{\gamma}} (\nu_0)^{\frac{1}{\gamma}} \sum_{k=1}^{N} \pi_k \left[\left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)\right]^{\frac{1}{\gamma}+1}. \quad (B.8)$$

Before isolating $(\nu_0)^{\frac{1}{\gamma}}$, it is interesting to see that

$$\sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1\right)^{\frac{1}{\gamma}+1} = \sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)^{\frac{1}{\gamma}}. \quad (B.9)$$

In fact, consider the first order condition in λ_i and μ_j

$$\frac{\partial}{\partial \lambda_i} \phi(\lambda, \mu, \nu_0) = -\frac{\gamma+1}{\gamma} \sum_{k=1}^{N} \pi_k \frac{1}{(\gamma+1)} \left[\gamma \nu_0 \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)\right]^{\frac{1}{\gamma}} \gamma \nu_0 A_{i,k} = 0 \quad (B.10)$$

$$\frac{\partial}{\partial \mu_j} \phi(\lambda, \mu, \nu_0) = -\frac{\gamma+1}{\gamma} \sum_{k=1}^{N} \pi_k \frac{1}{(\gamma+1)} \left[\gamma \nu_0 \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)\right]^{\frac{1}{\gamma}} \gamma \nu_0 B_{j,k} = 0 \quad (B.11)$$

and realize that

$$\sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)^{\frac{1}{\gamma}+1} = \sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)^{\frac{1}{\gamma}} \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)$$

$$= \sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)^{\frac{1}{\gamma}} + \sum_i \sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)^{\frac{1}{\gamma}} \lambda_i A_{i,k} +$$

$$\sum_j \sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)^{\frac{1}{\gamma}} \mu_j B_{j,k}$$

$$= \sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)^{\frac{1}{\gamma}}. \quad (B.12)$$

Finally, clearing $(\nu_0)^{\frac{1}{\gamma}}$

$$(\gamma \nu_0)^{\frac{1}{\gamma}} = \frac{1}{\sum_{k=1}^{N} \pi_k \left[\left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)\right]^{\frac{1}{\gamma}+1}} \quad (B.13)$$

$$= \frac{1}{\sum_{k=1}^{N} \pi_k \left[\left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1\right)\right]^{\frac{1}{\gamma}}}$$

$$\nu_0 = \frac{1}{\gamma} \left(\frac{1}{\sum_{k=1}^{N} \pi_k \left[\left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1 \right)^{\frac{1}{\gamma}} \right]} \right)^{\gamma}. \tag{B.14}$$

Now, remembering Eq. (B.2)

$$\overline{\pi}_k = \pi_k \left[\gamma \nu_0 \left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1 \right) \right]^{\frac{1}{\gamma}} = \pi_k \frac{\left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1 \right)^{\frac{1}{\gamma}}}{\sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} +_0 \sum_j \mu_j B_{j,k} + 1 \right)^{\frac{1}{\gamma}}}. \tag{B.15}$$

Appendix B.3. The dual problem

The dual problem in the main text is

$$\sup \left\{ -\frac{1}{(\gamma+1)} \sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1 \right)^{\frac{1+\gamma}{\gamma}} \right\}, \tag{B.16}$$

and the first order conditions are

$$\sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1 \right)^{\frac{1}{\gamma}} A_{i,k} = 0, \tag{B.17}$$

$$\sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1 \right)^{\frac{1}{\gamma}} B_{j,k} = 0. \tag{B.18}$$

The dual problem derived here is

$$\sup \left\{ -\frac{1}{\gamma(\gamma+1)} + \nu_0 - \sum_{k=1}^{N} \pi_k \frac{1}{(\gamma+1)} \left[\gamma \nu_0 \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1 \right) \right]^{\frac{1}{\gamma}+1} \right\}, \tag{B.19}$$

and the first order conditions are

$$\sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1 \right)^{\frac{1}{\gamma}} A_{i,k} = 0, \tag{B.20}$$

$$\sum_{k=1}^{N} \pi_k \left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1 \right)^{\frac{1}{\gamma}} B_{j,k} = 0,$$

$$\nu_0 = \frac{1}{\gamma} \left(\frac{1}{\sum_{k=1}^{N} \pi_k \left[\left(\sum_i \lambda_i A_{i,k} + \sum_j \mu_j B_{j,k} + 1 \right)^{\frac{1}{\gamma}} \right]} \right)^{\gamma}. \tag{B.21}$$

Note that the optimal λ_i and μ_j define the optimal ν_0. Moreover, the first order condition in both problems coincides. When the objective function is convex and differentiable, the first order condition determines the optimal variables.