The Minimum-Entropy Algorithm and Related Methods for Calibrating Asset-Pricing Models

Marco Avellaneda

Abstract. We describe an algorithm for calibrating asset-pricing models based on minimizing the relative entropy between probabilities. The algorithm determines a probability measure on path-space which minimizes the Kullback information with respect to a given prior and satisfies a finite number of moment constraints which correspond to fitting prices. It admits, generically, a unique, stable, solution that depends smoothly on the input prices. We study the sensitivities of the model values of contingent claims to variations in the input prices. We find that hedge ratios can be interpreted as "risk-neutral" regression coefficients of the contingent claim's payoff on the set of payoffs of the input instruments. We also show that the minimum-entropy algorithm is a special case of a general class of algorithms for calibrating asset-pricing models based on stochastic control and convex optimization. As an illustration, we use minimum-entropy to construct a smooth curve of instantaneous forward rates from US LIBOR data and to study the corresponding sensitivities of fixed-income securities to variations in input prices.

1 Introduction

Despite its practical importance, model calibration has received little attention in Mathematical Finance. Calibrating an asset-pricing model means specifying a probability distribution for the underlying state-variables in such a way that the model reproduces, by taking discounted expectations, the current market prices of a set of reference securities. The reference securities, or inputs, characterize the market under consideration. The most common models of this kind are yield-curve models, used for managing portfolios of fixed-income securities.[1] Other, less ubiquitous, examples are the so-called local volatility models used for managing option portfolios.[2]

[1]In this case, it is customary to vary the swap rates or bond yields corresponding to standard maturities by one basis point and to compute the corresponding dollar change in the portfolio value. These sensitivities are the so-called "DV01"s (dollar value of one basis point) used to quantify interest-rate exposure.

[2]Also known as "smile models".

In many cases of interest, the calibration problem is equivalent to a classical problem in statistics: the determination of a probability distribution from a finite set of moments. The "moments" correspond to the discounted expectations of the cash-flows of the reference instruments. It is well-known, however, that such problems are ill-posed: there can be many solutions or, sometimes, no solution at all. In financial-economic terms, this signifies that prices may not be consistent with *any* risk-neutral probability (and hence that an arbitrage exists) or, more likely, that there exist several risk-neutral probabilities consistent with the current prices due to market incompleteness. Selecting a probability is tantamount to "completing the market", in the sense that Arrow-Debreu prices are assigned to all future states. Thus, any calibration procedure involves making subjective choices. Taking into account available econometric information and stylized facts about the market reduces (partially) the ill-posedness of the model selection problem. Intuitively, a calibrated model which is "near" our prior beliefs and market knowledge is more desirable than one that is "far away" from the prior. [3]

In this paper, we study an algorithm which consists in choosing the risk-neutral probability that minimizes the *relative entropy*, or *Kullback-Leibler entropy* with respect to a subjective prior. This approach was pioneered in statistics by Jaynes (1996) and others; see McLaughlin (1984), Cover and Thomas (1991). An appealing feature of the method is that it takes into account the *a priori* (econometric) information available. This information is modeled by the prior probability, which can be viewed as a "first step" towards adjusting the model to econometric data but not necessarily to current prices. The entropy minimization algorithm provides a way of reconciling the prior with the information contained in current market prices.

Buchen and Kelly(1996) and Gulko(1995, 1996) used entropy minimization for calibrating one-period asset pricing models; see also Jackwerth and Rubinstein (1996) and Platen and Rebolledo (1996). In a previous article, Avellaneda, Friedman, Holmes and Samperi (1997) applied the minimum relative entropy method to the calibration of volatility surfaces in the context of commodity option pricing. In the present paper, following Buchen and Kelly and Avellaneda *et al*, we use Lagrange multipliers to model the price constraints. However, we go one step further in the analysis and study also the sensitivities of the model with respect to the input prices. For this purpose, we use the matrix of second derivatives with respect to the Lagrange multipliers computed at the critical point.

The paper is organized as follows: In Section 2, we consider a one-period model. Under mild no-degeneracy assumptions, we show that if there exists a probability with finite relative entropy, then the calibration problem has a unique solution. We establish also that the price-sensitivities of contingent claims depend smoothly on the input prices. The calibrated model has a remarkable property: the *deltas* (price-sensitivities) and the *betas* (regression coefficients of the cash-flow of a contingent claim on the space generated by the cash-flows of the input instruments)

[3]For example, practitioners tend to favor models in which interest rates are mean-reverting and oscillate about some asymptotic distribution. Processes that have unit roots and can reach very large values with large probabilities are discarded and appear to fail to pass simple statistical tests.

are, in fact, equal. More precisely, let Π denote the model price of a contingent claim which has a discounted payoff h. Let us denote by $G_i,\ i = 1, 2, ..., N$ the discounted cash-flows of the reference instruments, and by $C_1, \ldots C_N$ their prices. Then, we have

$$\frac{\partial \Pi}{\partial C_i} = \sum_{j=1}^{N} K_{ij} \operatorname{Cov}\{G_j, h\}$$

where

$$K = H^{-1} \quad , \quad H_{ij} = \operatorname{Cov}\{G_i, G_j\}$$

and COV represents the covariance operator under the risk-neutral (calibrated) measure. It is well-known that the right-hand side of the first equation corresponds to the value of the coefficient β_i in the linear regression model

$$h = \alpha + \sum_{i=1}^{N} \beta_i G_i + \epsilon$$

where ϵ has mean zero and is uncorrelated with the cash-flows $\{G_j\}$ under the risk-neutral measure. This property of the minimum-entropy algorithm suggests that it has econometric relevance. [4]

Sections 3 and 4 are devoted to inter-temporal asset-pricing models, where we formulate the algorithm in terms of partial differential equations. The algorithm involves solving a Hamilton-Jacobi-Bellman partial differential equation of "quasi-linear" type[5] and minimizing the value of the solution at one point in terms of a finite set of Lagrange multipliers. The gradient of the objective function corresponds to a coupled system of linearized equations.

In Section 5, we show that the algorithm can be formulated as a constrained stochastic control problem. This suggests that there are many generalizations of the "pure" entropy algorithm that can be made by changing the form of the cost function. Specifically, minimizing relative entropy is equivalent to minimizing the L_2 norm of the risk-premia $m_i(t)$, *i.e.*

$$\mathrm{E}^P \left\{ \int_0^{T_{max}} \sum_{i=1}^{\nu} m_i(t)^2 \, dt \right\}$$

where T_{max} is the time-horizon and ν is the number of factors. In practice, it is computationally advantageous to consider functionals of the form

$$\mathrm{E}^P \left\{ \int_0^{T_{max}} e^{-\int_0^t r(s)\, ds} \sum_{i=1}^{\nu} m_i(t)^2 \, dt \right\} ,$$

[4] Calibration via relative entropy minimization is, in a certain sense, the non-parametric counterpart of the maximum-likelihood estimation method; cf Jaynes (1996).

[5] This means that the nonlinearity appears in the gradient terms.

because this reduces the dimensionality of the computation, while preserving at the same time the essential features of the algorithm. [6]

In Section 6, we use the algorithm to construct smooth forward rate curves from US LIBOR data (FRAs and swap rates). We pay particular attention to the sensitivities with respect to input swap rates, an issue that remains somewhat controversial among practitioners. Hedges tend to be model-dependent and therefore a certain amount of risk is taken when choosing different forward rate curves. The issue is whether smooth curves, which give rise to "non-local" hedges[7], are preferable to discontinuous forward rate curves, such as the ones obtained by the bootstrapping method. The latter method tends to give rise to "local" hedges in which the sensitivities are essentially limited to the nearest maturities. It is our hope that the minimum-entropy method can compete favorably and perhaps even improve on some of the other methods used to generate smooth forward-rate curves, in the sense that the resulting sensitivities are acceptable from a practical viewpoint. These issues will be investigated in a separate paper.

2 Relative entropy minimization with moment constraints

We consider the problem of determining a probability density function $f(X)$ for a real-valued random variable X satisfying

$$\int G_i(X)\, f(X)\, dX \;=\; C_i\,, \quad 1 \;\le\; i \;\le\; N\;, \tag{1}$$

where $G_1(X)$, $...G_N(X)$ are given functions and C_1, $...C_N$ are given numbers. [8] Financially, X represents a state-variable describing the economy; $G_i(X)$ and C_i represent, respectively, the cash-flows and prices of a set of traded securities.

Buchen and Kelly proposed, in the context of option pricing, to choose the density $f(X)$ that minimizes the functional

$$H(f|f_0) \;=\; \int f\, log\left(\frac{f}{f_0}\right)\; dX\;, \tag{2}$$

where $f_0(X)$ is a prior probability density function. The expression $H(f|f_0)$ is known as the Kullback-Leibler entropy or relative entropy of f with respect to f_0. It represents the "information distance" between $f(X)$ and $f_0(X)$. [9]

It is well-known (Cover and Thomas) that if there exists a probability density function f satisfying the constraints (1) and such that $H(f|f_0)$ is finite, the solution of the constrained entropy minimization problem exists and can be found by the method of Lagrange multipliers. Namely, we solve

[6] The advantage of passing from minimum-entropy to a more general control formulation was also shown in Avellaneda *et. al.*, where the technique was used to "regularize" the relative entropy of two mutually singular diffusions.

[7] By this we mean hedges that imply correlations between bonds with distant maturities.

[8] Henceforth, we say that a probability satisfying the constraints (1) is *calibrated*. It is implicitly assumed that the functions $G_i(X)$ are such that all integrals considered are well-defined.

[9] The relative entropy is not symmetric with respect to the variables f and f_0, so it is not a distance in the mathematical sense of the word. Nevertheless, it measures the "deviation" of f from f_0 (Cover and Thomas).

$$\underset{\lambda_i}{inf}\ \underset{f}{sup}\ \left[-H(f|f_0)\ +\ \sum_i \lambda_i\ \left(\int G_i\, f\, dX\ -\ C_i \right) \right] . \tag{3}$$

Let us first fix λ and seek the density that maximizes this "augmented Lagrangian". An elementary calculation of the first-order optimality conditions (Cover and Thomas) shows that for each λ, the optimal probability density function is given by

$$f_\lambda(X) \quad = \quad \frac{1}{Z(\lambda)}\ f_0(X)\, e^{\sum_i \lambda_i\, G_i(X)} \tag{4}$$

where $Z(\lambda)$ is the normalization factor

$$Z(\lambda) \quad = \quad \int f_0\ e^{\sum_i \lambda_i\, G_i}\ dX.$$

Substituting expression (3a) into (2), it follows that the optimization over the Lagrange multipliers is equivalent to minimizing the function

$$\log\left(Z(\lambda)\right)\ -\ \sum_i \lambda_i\, C_i\ , \tag{5}$$

over all values of $\lambda\ =\ (\lambda_1,\ ...\ \lambda_N)$. The first-order conditions for a minimum are

$$\frac{1}{Z(\lambda)}\ \frac{\partial Z(\lambda)}{\partial \lambda_i}\ =\ C_i\ .$$

This shows, in view of (4), that if λ is a critical point of (5) then f_λ is calibrated.

The stability of the solution, i.e. the continuous dependence of f_λ on input prices, follows from convex duality. To see this, notice first that

$$\begin{aligned} \left(log\left(Z(\lambda)\right)\right)_{\lambda_i,\,\lambda_j} \quad &= \quad \frac{Z_{\lambda_i\ \lambda_j}}{Z}\ -\ \frac{Z_{\lambda_i}\, Z_{\lambda_j}}{Z^2} \\ &= \quad \mathrm{Cov}_{f_\lambda}\ [G_i(X),\ G_j(X)\ ]\ \equiv\ H_{ij}\ . \end{aligned}$$

Since covariance matrices are non-negative definite, $log(Z(\lambda))\ -\ \lambda \cdot C$ is convex. It also follows from this characterization that $log(Z(\lambda))$ is strictly convex if the N payoff functions are linearly independent. [10]

Let λ^* be the value of the Lagrange multipliers that minimizes the objective function $log\,[Z(\lambda)]\ -\ \lambda\, C$. To assess the sensitivity of the calibrated probability

[10]As a rule, redundancies within the class of input securities should be avoided when fitting prices. They lead to instabilities, since the input prices must satisfy linear relation exactly (i.e. with infinite precision) in order to avoid mispricing these instruments with the model.

f_{λ^*} to input prices, consider a new contingent claim with payoff $h(X)$ (the "target payoff"). Let $\Pi(\lambda) \;=\; \mathrm{E}^{f_\lambda}\,(h(X))$. Then, we have

$$\begin{aligned}
\frac{\partial \Pi(\lambda^*)}{\partial \lambda_j} &= \frac{\partial}{\partial \lambda_j}\,\frac{\int f_0\, e^{\lambda\cdot G}\, h\, dX}{\int f_0\, e^{\lambda\cdot G}\, dX} \\
&= \mathrm{E}^{f_\lambda}(h(X)\, G_j(X)) \;-\; \mathrm{E}^{f_\lambda}(h(X))\; \mathrm{E}^{f_\lambda}(G_j(X)) \\
&= \mathrm{Cov}_{f_{\lambda^*}}\,(\; h(X),\; G_j(X))\;\; .
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\partial \Pi(\lambda^*)}{\partial C_i} &= \sum_j \left(\frac{\partial \Pi(\lambda)}{\partial \lambda_j}\right)_{\lambda\,=\,\lambda^*} \frac{\partial \lambda_j^*}{\partial C_i} \\
&= \sum_j \left(\frac{\partial \Pi(\lambda^*)}{\partial \lambda_j}\right)_{\lambda\,=\,\lambda^*} \left(H^{-1}\right)_{ij} \\
&= \sum_j \mathrm{Cov}_{f_{\lambda^*}}\,(\; h(X),\; G_j(X))\,\left(H^{-1}\right)_{ij}\;\; . \qquad (6)
\end{aligned}$$

Here, in deriving the second equation, we made use of the well-known duality relations (Rockafellar (1970))

$$\frac{\partial C_i}{\partial \lambda_j^*} \;=\; H_{ij} \qquad , \qquad \frac{\partial \lambda_j^*}{\partial C_i} \;=\; \left(H^{-1}\right)_{ij}\;\; .$$

It follows from equations (4) and (6) that $\Pi \;=\; \Pi\,(C_1,\, ...C_N)$ is infinitely differentiable as a function of $C_1,\, ...\, C_N$. In particular the sensitivities $\frac{\partial \Pi}{\partial C_i}$ vary continuously with the input prices.

Formula (6) admits a simple interpretation. Consider the linear regression model

$$h(X) \;=\; \alpha \;+\; \sum_{i=1}^{N} \beta_i\; G_i(X) \;+\; \epsilon\;,$$

where we assume that ϵ is a random variable with mean zero uncorrelated with $G_i(X)$ $i \;=\; 1,\, ..N$ under the the risk-neutral measure. The coefficients β_i which minimize the variance of the residual $h \;-\; \alpha \;-\; \sum_i \beta_i\, G_i$ are given by

$$\beta_i \;=\; \sum_j \left(H^{-1}\right)_{ij}\; \mathrm{Cov}_{f_{\lambda^*}}\,(h(X),\, G_j(X)) \;=\; \frac{\partial \Pi}{\partial C_i}\;, \quad 1 \;\le\; i \;\le\; N\;.$$

We summarize the results of this section in

PROPOSITION 1. *(a) The minimum-relative-entropy method reduces the class of candidate solutions of the moment problem to an N-parameter exponential family $f_\lambda(X)$ given by (4).*

Assume that the input payoffs $G_1(X)$, ... $G_N(X)$ are linearly independent. Then:

(b) If there exists a calibrated density $f(X)$ such that $H(f|f_0) < \infty$, the solution of the constrained entropy-minimization problem is unique.

(c) The sensitivities of contingent-claim prices to variations in input prices are equal to the linear regression coefficients of the target payoff on the input payoffs under the calibrated measure.

3 INTER-TEMPORAL MODELS

We consider a classical continuous-time economy, represented by a state-vector $\mathbf{X}(t) = (X_1(t), ..., X_\nu(t))$ which follows a diffusion process under the prior probability measure:

$$dX_i(t) = \sum_{j=1}^{\nu} \sigma_{ij}^{(0)} \, dZ_j(t) + \mu_i^{(0)} \, dt \qquad 1 \leq i \leq \nu . \tag{7}$$

Here $(Z_1, ... Z_\nu)$ are independent Brownian motions and $\sigma_{ij}^{(0)}$ and $\mu^{(0)}$ are functions of X and t.

We assume that there are N traded securities, with prices C_1, ... C_N. Our goal is to find a risk-neutral probability measure P consistent with these prices based on the principle of minimum relative entropy with respect to the prior (denoted by P_0).

The price constraints can be written in the form on N equations

$$C_i = \mathrm{E}^P \left\{ \sum_{k=1}^{n_i} e^{-\int^{T_{ik}} r(s)\, ds} \, G_{ik}(\mathrm{X}(T_{ik})) \right\} , \qquad 1 \leq i \leq N , \tag{8}$$

where $\{T_{ik}\}_{k=1}^{n_i}$ are the cash-flow dates of the i^{th} security and $\{G_{ik}(\mathrm{X})\}_{j=1}^{n_i}$ represent the corresponding cash-flows. We assume that the latter are bounded, continuous functions of X. The process $r(s) = r(\mathrm{X}(s), s)$ represents the short-term (continuously compounded) interest rate. Notice that in (8) the expectation value is taken with respect to a calibrated (risk-neutral) measure P which, in general, is not equal to P_0.

We follow the approach of the previous section. First, we consider the Kullback-Leibler relative entropy of P with respect to P_0 in the diffusion setting. For this purpose, it is convenient to define a finite time horizon $0 < t < T_{max}$, (where $T_{max} \geq \max_{ik} T_{ik}$). The relative entropy of P with respect to P_0 is given by

$$H(P|P_0) = \mathrm{E}^P \left\{ \log \left(\frac{dP}{dP_0} \right)_{T_{max}} \right\} ,$$

where $\left(\frac{dP}{dP_0}\right)_{T_{max}}$ is the Radon-Nykodym derivative of P with respect to P_0 over the time-horizon T_{max} . [11]

Next, we consider the augmented Lagrangian associated with the constraints (8) (compare with (3))

$$-\mathrm{E}^P\left\{\log\left(\frac{dP}{dP_0}\right)_{T_{max}}\right\}+\sum_{i=1}^{N}\lambda_i\left(\sum_j \mathrm{E}^P\left\{e^{-\int_0^{T_{ij}} r(s)ds} G_{ij}(\mathrm{X}(T_{ij}))\right\}-C_i\right). \quad (9)$$

The solution of the inf-sup problem is identical to the one outlined in the previous section. Accordingly, we define the normalization factor (cf. (4))

$$Z(\lambda) \;=\; \mathrm{E}^{P_0}\left\{\exp\left(\sum_{ij}\lambda_i\, e^{-\int_0^{T_{ij}} r(s)ds} G_{ij}(\mathrm{X}(T_{ij}))\right)\right\}. \quad (10)$$

Further, by mimicking equation (4), we obtain a parametric family of measures $\{P_\lambda\}_\lambda$ defined by their Radon-Nykodym derivatives with respect P_0:

$$\frac{dP_\lambda}{dP_0} \;=\; \frac{1}{Z(\lambda)}\cdot\exp\left(\sum_{ij}\lambda_i\, e^{-\int_0^{T_{ij}} r(s)ds} G_{ij}(\mathrm{X}(T_{ij}))\right). \quad (11)$$

Elementary calculus of variations shows that for any fixed vector $\lambda = (\lambda_1,\, \lambda_2,\, ...\lambda_N)$, the measure P_λ realizes the supremum of the Lagrangian (9) over all probability measures. As expected, the supremum is given by

$$log[\,Z(\lambda)\,] \;-\; \sum_{i=1}^{N}\lambda_i\, C_i\;.$$

If λ is a critical point, we have

$$C_i \;=\; \frac{Z_{\lambda_i}}{Z} \;=\; \mathrm{E}^{P_\lambda}\left\{\sum_{j=1}^{n_i} e^{-\int_0^{T_{ij}} r(s)ds} G_{ij}(\mathrm{X}(T_{ij}))\right\} \qquad 1\,\le\, i\,\le\, N\;.$$

Therefore, the corresponding measure P_λ is calibrated to the input prices.

[11] In particular, the relative entropy is infinite if P is not absolutely continuous with respect to P_0.

Define the discounted cash-flows

$$\Gamma_i \;=\; \sum_{j=1}^{n_i} e^{-\int_0^{T_{ij}} r(s)ds}\, G_{ij}(\mathrm{X}(T_{ij}))\;, \quad 1 \;\leq\; i \;\leq\; N\;.$$

As in the previous section, we can interpret the Hessian of $log(Z(\lambda)) - \lambda C$ as a covariance matrix, *viz.*,

$$\frac{\partial^2}{\partial\lambda_i\,\partial\lambda_j}\,(log(Z(\lambda)) - \lambda\,) \;=\; \mathrm{Cov}^P\,[\Gamma_i,\;\Gamma_j]\;.$$

Similarly, if $h(X_T)$ is the payoff of a security maturing at time $T \;\leq\; T_{max}$, we have

$$\frac{\partial}{\partial\lambda_j}\mathrm{E}^P\left\{e^{-\int_0^T r_s\,ds}\,h(X_T)\right\} \;=\; \mathrm{Cov}^P\left[\Gamma_j,\; e^{-\int_0^T r_s\,ds}\,h(X_T)\right]\;.$$

Like in the previous section, we conclude that

Proposition 2. *(a) Relative entropy minimization is equivalent assuming that the probability measure belongs to an N-parameter exponential family given by (11).*

(b) If the input payoffs are linearly independent, there is at most one calibrated measure that minimizes relative entropy.

(c) The model prices and sensitivities of contingent claims depend continuously on input prices.

(d) The sensitivities with respect to input prices can be interpreted as the linear regression coefficients of the target discounted cash-flows on the space generated by the discounted cash-flows of the input instruments.

4 PDE formulation

Let $\mathrm{L}^{(0)}$ represent the infinitesimal generator of the semi-group corresponding to the prior P_0 *i.e.*,[12]

$$\mathrm{L}^{(0)}\,\phi \;=\; \frac{1}{2}\sum_{ij} a_{ij}\,\phi_{X_i\,X_j} \;+\; \sum_i \mu_i^{(0)}\,\phi_{X_i} \qquad (12)$$

where

$$a_{ij} \;=\; \sum_{p=1}^{\nu} \sigma_{ip}^{(0)}\,\sigma_{jp}^{(0)}\;.$$

[12]We use the notation $\phi_{X_i} \;=\; \frac{\partial\phi}{\partial X_i}$ for partial derivatives.

It follows from (10) and standard diffusion theory that the normalization factor is given by

$$Z(\lambda) \;=\; U(\mathrm{X}(0),\,1,\,0\,;\lambda)\;, \tag{13}$$

where $U(\mathrm{X},\,Y,\,t\,;\lambda)$ is the solution of the Cauchy problem

$$U_t \;+\; \mathrm{L}^{(0)}\,U \;-\; r\,Y\,U_Y \;=\; 0 \qquad t \;\neq\; T_{ij} \tag{14}$$

with the boundary conditions at cash-flow dates $t = T_{ij}$

$$U(\mathrm{X},\,Y,\,T_{ij}-0\,;\lambda) \;=\; U(\mathrm{X},\,Y,\,T_{ij}+0\,;\lambda)\cdot\exp\left(\sum_i \lambda_i\,G_{ij}(\mathrm{X})\,Y\right)\;. \tag{15}$$

To derive (14), we introduced the auxiliary state-variable $Y_t \;=\; e^{-\int_0^t r(s)\,ds}$ and the $\nu+1$-dimensional process $(\mathrm{X}_t,\,Y_t)$ which is a Markov process with an infinitesimal generator given by the left-hand side of (14).

From (14) we can derive partial differential equations satisfied by $log(Z(\lambda))$ and its gradient with respect to λ. Accordingly, we obtain

$$log(Z(\lambda)) \;=\; W(\mathrm{X}(0),\,1,\,0\,;\lambda)\;, \qquad \frac{Z_{\lambda_i}}{Z} \;=\; V^{(i)}(\mathrm{X}(0),\,1,\,0\,;\lambda)$$

where W satisfies the PDE

$$W_t + \mathrm{L}^{(0)}W + \frac{1}{2}\sum_{ij=1}^{N} a_{ij}\,W_{X_i}\,W_{X_j} - r\,Y\,W_Y \;=\sum_{ij}\lambda_i\,G_{ij}(\mathrm{X})\,Y\,\delta(t-T_{ij})\,. \tag{16}$$

The PDE for $V^{(l)}$ is obtained by differentiating (16) with respect to λ_l, *viz.*

$$V_t^{(l)} \;+\; \mathrm{L}^{(0)}V^{(l)} \;+\; \sum_{ij=1}^{N} a_{ij}\,W_{X_i}\,V_{X_j}^{(l)} \;-\; r\,Y\,V_Y^{(l)} \;=$$

$$\sum_{j=i}^{n_i} G_{ij}(\mathrm{X})\,Y\,\delta(t-T_{ij})\;. \tag{17}$$

From this last equation, we deduce the following characterization of the calibrated measure.

Proposition 3. *The calibrated measure which minimizes the relative entropy corresponds to the diffusion process*

$$dX_i \;=\; \sum_{j=1}^{\nu}\sigma_{ij}^{(0)}\;dZ_j \;+\; \left(\;\mu_i^{(0)} \;+\; \sum_{j=1}^{\nu}\sigma_{ij}^{(0)}\,m_j\right)\;dt$$

with

$$m_i \;=\; \sum_{i=1}^{N} \sigma_{ij}^{(0)}\, W_{X_j} \;, \tag{18}$$

where W is computed with λ at the critical point.

5 Modified entropies and the optimal control formulation.

It is useful to view the entropy minimization algorithm as a stochastic optimal control problem with constraints. We recall the following result (Platen and Rebolledo): Proposition 4. *The class of diffusion measures P which have finite*

relative entropy with respect to P_0 consists of Ito processes

$$dX_i(t) \;=\; \sum_{j=1}^{\nu} \sigma_{ij}^{(0)}\, dZ_j(t) \;+\; \mu_i\, dt$$

with

$$\mu_i \;=\; \mu_i^{(0)} \;+\; \sum_{j} \sigma_{ij}^{(0)}\, m_j \;,$$

where, $m_j \quad 1 \;\le\; j \;\le\; \nu$ are square-integrable. Moreover, the relative entropy of P with respect to P_0 (viewed as measures in path-space with the time horizon $0 \;<\; t \;<\; T_{max} \;=\; \max\limits_{ik} T_{ik}$) is given by

$$H(P|P_0) \;=\; \frac{1}{2}\,\mathrm{E}^P \left\{ \int\limits_0^{T_{max}} \sum_{j=1}^{\nu} m_j(t)^2\, dt \right\} \;. \tag{19}$$

Thus, minimizing the KL entropy is equivalent to selecting the risk-neutral measure in such a way that the vector of risk-premia has the smallest mean-square norm (cf. Platen and Rebolledo (1996), Samperi(1997)).

Using (19) we rewrite the augmented Lagrangian (9) as

$$-\mathrm{E}^P \left[\int\limits_0^{T_{max}} \sum_{j=1}^{\nu} m_j^2(t)\, dt \right] \;+\; \sum_{i=1}^{N} \lambda_i\, \mathrm{E}^P \left[\sum_{j=1}^{n_i} e^{-\int\limits_0^{T_{ij}} r(s)ds}\, G_{ij}(\mathrm{X}(T_{ij})) \right] \tag{20}$$

The advantage of the stochastic control formulation is that it can be generalized considerably. In fact, we can replace the function $\sum\limits_j m_j^2(t)$ by more general functions of the form $\eta(_t,\, m_1(t),\, m_2(t),\, ...m_\nu(t))$, which are strictly convex in $m_i(t)$.

The class of functionals of the form

$$H_{mod}(P|P_0) \;=\; \frac{1}{2}\,\mathrm{E}\left\{ \int_0^{T_{max}} e^{-\int_0^t r(s)\,ds}\quad \eta(m(t))\;\, dt \right\}\;, \tag{21}$$

where $\eta(m)$ is a deterministic and strictly convex is of particular importance. In this case, $H_{mod}(P|P_0)$ can be seen as a "running cost" with respect to the choice of parameters which penalizes deviations from the prior.

Notice that the definition of entropy in (19) is independent of the interest rate. One important advantage of discounting the local entropy by the interest rate is *dimension reduction*: we can dispense of the auxiliary state variable Y. In fact, the HJB equation corresponding to the modified entropy (21) is

$$W_t \;+\; \mathrm{L}^{(0)}W \;+\; \eta^*\left(\sigma^{(0)}\cdot W_X\right) \;-\; r\,W \;=$$

$$\sum_{ij} \lambda_i\, G_{ij}(\mathrm{X})\,\delta(t-T_{ij})\;, \tag{22}$$

where η^* is the Legendre transform of η (Rockafellar). The function W plays the role of $log(Z(\lambda)$ in the "pure entropy" framework. Note, however, that in the special case $\eta(t,\, m) \;=\; \frac{1}{2}\sum_j m_j^2$ we have $\eta = \eta^*$. The corresponding Bellman equation is

$$W_t \;+\; \mathrm{L}^{(0)}W \;+\; \frac{1}{2}\sum a_{ij}\,W_{X_i}\,W_{X_j} \;-\; r\,W \;=$$

$$\sum_{ij} \lambda_i\, G_{ij}(\mathrm{X})\,\delta(t-T_{ij})\;, \tag{23}$$

In the rest of this section we assume this particular form for the modified entropy. Following the steps outlined in §2, the algorithm consists of minimizing

$$W\,(\mathrm{X}(0),\,0\,;\;\lambda_1,\;\ldots\,\lambda_N) - \sum_{i=0}^{M}\lambda_i\,C_i\;,$$

over λ. This is done with a gradient-based optimization algorithm such as L-BFGS (Zhu, Boyd, Lu and Nocedal (1994)). The gradient is computed by solving the N linearized equations:

$$V_t^{(l)} \;+\; \mathrm{L}^{(0)}\,V^{(l)} \;+\; \sum_{ij} a_{ij}\,W_{X_i}\,V_{X_j}^{(l)} \;-\; r\,V^{(l)} \;=$$

$$\sum_{j=0}^{n_l} G_{lj}(\mathrm{X})\,\delta(t-T_{lj}) \tag{24}$$

Notice that the first-order conditions for the minimum in λ are

$$V^{(l)}\left(\mathrm{X}(0),\,0\,;\;\lambda_1,\;...\;\lambda_N\right)\;-\;\lambda_i\,C_l\;=\;0\;,\qquad 1\;\leq\;l\;\leq\;N\;\;.$$

Formally, these equations imply that the corresponding probability measure is calibrated, since

$$V^{(l)}\left(\mathrm{X}(0),\,0\,;\;\lambda_1,\;...\;\lambda_N\right)\;=\;\mathrm{E}^P\left\{\sum_{k=1}^{n_l}\;e^{-\int^{T_{lk}} r(s)\,ds}\,G_{lk}(\mathrm{X}(T_{lk})\right\}\;.$$

Here P is the diffusion process with drift $\mu^{(0)}+W_X\cdot\sigma^{(0)}$, where W is calculated at the optimal values of the Lagrange multipliers. We refer to the diffusion measure implied by solving equation (23) as P_λ, a slight abuse of notation. The optimal control formulation has the same mathematical structure (i.e. convexity λ) as the "pure" entropy problem. To study the dependence on the inputs, we consider the Hessian of $W(\lambda)$. Differentiating equations (24) with respect to λ, we find that the Hessian matrix

$$H^{(lm)}\;=\;\frac{\partial^2 W}{\partial\lambda_l\partial\lambda_m}$$

satisfies

$$H_t^{(lm)}\;+\;L\,H^{(lm)}+\sum_{ij}a_{ij}\,W_{X_i}\,H_{X_j}^{(lm)}\;+\sum_{ij}a_{ij}\,V_{X_i}^{(l)}\,V_{X_j}^{(m)}\;-\;r\,H^{(lm)}\;\;=0\;.\quad(25)$$

In particular, we have

$$H^{(lm)}(\mathrm{X}(0),\,0;\,\lambda^*)\;=\;\mathrm{E}^P\left\{\int_0^{T_{max}}e^{-\int_0^t r(s)\,ds}\;\sum_{i\,j\,=\,1}^{M}a_{ij}\,V_{X_i}^{(l)}\,V_{X_j}^{(m)}dt\right\}\;.\qquad(26)$$

Unlike the case of "pure" entropy, the Hessian does not admit a simple interpretation in terms of linear regression coefficients. Nevertheless, we can express the difference between the Hessian and the covariance matrix of the discounted input cash-flows as an expectation. More precisely, we have

$$\mathrm{Cov}^{P_\lambda}\left(\Gamma^{(l)},\;\Gamma^{(m)}\right)\;=\;\mathrm{E}^{P_\lambda}\left\{\int_0^{T_{max}}e^{-2\int_0^t r(s)\,ds}\;\sum_{ij=1}^{\nu}a_{ij}\,V_{X_i}^{(l)}\,V_{X_j}^{(m)}\;dt\right\}\;,\;(27)$$

which differs from (26) in the fact that the stochastic discount factor is squared. Therefore, we conclude that

$$H^{(lm)} \;=\; \mathrm{Cov}^{P_\lambda}\left(\Gamma^{(l)},\; \Gamma^{(m)}\right) \;+$$

$$\mathrm{E}^{P_\lambda}\left\{\int_0^{T_{max}}\left(e^{-\int_0^t r(s)\,ds} \;-\; (e^{-2\int_0^t r(s)\,ds}\right)\sum_{ij=1}^{\nu} a_{ij}\,V^{(l)}_{X_i}\,V^{(m)}_{X_j}\;dt\right\}\,. \tag{28}$$

In particular, this shows that if the instruments are not linearly dependent with P_0-probability 1, the Hessian matrix is positive definite.[13] [14] Barring trivial redundancies, the argument establishes that there is at most one λ that minimizes the objective function.

Finally, we analyze the sensitivities of model prices to input prices.

Given a contingent claim with a payoff $h(X_T)$ due date T, $(T \,<\, T_{max})$, let Π and $\Pi^{(l)}$ denote, respectively, the model price and the sensitivity of this price with respect to λ_l.

The functions Π and $\Pi^{(l)}$ are readily computed by solving the system of equations

$$\Pi_t \;+\; \mathrm{L}^{(0)}\,\Pi \;+\; \sum_{ij} a_{ij}\,W_{X_i}\,\Pi_{X_j} \;-\; r\,\Pi \;=\delta(t-T)\,h(X)\;, \tag{29}$$

and

$$\Pi^{(l)}_t \;+\; \mathrm{L}^{(0)}\,\Pi^{(l)} \;+\; \sum_{ij} a_{ij}\,W_{X_i}\,\Pi^{(l)}_{X_j} \;+\; \sum_{ij} a_{ij}\,\Pi_{X_i}\,V^{(l)}_{X_j} \;-\; r\,\Pi^{(l)} \;=0\,. \tag{30}$$

It follows from this that the $\Pi^{(l)} \;=\; \Pi^{(l)}(\mathrm{X}(0),\,0)$ satisfies

$$\begin{aligned}\Pi^{(l)} \;\;&=\;\; \mathrm{E}^{P_\lambda}\left\{\int_0^{T_{max}} e^{-\int_0^t r(s)\,ds}\;\sum_{ij=1}^{\nu} a_{ij}\,V^{(l)}_{X_i}\,\Pi_{X_j}\;dt\right\}\\ &=\;\; \mathrm{Cov}^{P_\lambda}\left[e^{-\int_0^T r(s)\,ds}\;h(\mathrm{X}_T),\;\;\Gamma^{(l)}\right] \;+\end{aligned}$$

[13]This property also follows directly from equation (25). The strict positivity of the Hessian holds for any strictly convex modified entropy function $\eta(\mathrm{M},\,t)$, provided that the inputs are not linearly dependent.

[14]For example, the following set of inputs is linearly dependent, or redundant: (i) a one-year swap resetting quarterly, and (ii) four 3-month forward-rate agreements starting at the swap reset dates. This constitutes a redundancy because the swap can be replicated exactly with the FRAs.

$$\mathrm{E}^{P_\lambda}\left\{\int_0^{T_{max}}\left(e^{-\int_0^t r(s)\,ds} - e^{-2\int_0^t r(s)\,ds}\right)\sum_{ij=1}^{\nu} a_{ij}\,V_{X_i}^{(l)}\,\Pi_{X_j}\;dt\right\}. \tag{31}$$

As in §2, we can compute the sensitivities of Π with respect to the input prices C_1, ..., C_N using the inverse Hessian and the sensitivities with respect to λ. Accordingly, we have

$$\begin{aligned}\frac{\partial \Pi}{\partial C_m} &= \sum_{l=1}^{N}\frac{\partial \Pi}{\partial \lambda_l}\frac{\partial \lambda_l}{\partial C_m}\\ &= \sum_{l=1}^{N}\Pi^{(l)}\;\left(H^{-1}\right)_{lm}\end{aligned} \tag{32}$$

where H^{-1} is the inverse of H.

6 Forward-rate modeling and hedging portfolios of interest rate swaps

To illustrate the minimum-entropy algorithm, we calibrate a one-factor interest-rate model to the prices of standard instruments in the US LIBOR market.

We consider a set of input instruments consisting of forward-rate agreements (FRAs) and swaps with standard maturities. Using the algorithm, we compute a probability measure on the process driving the short-term rate which has the property that all the input instruments are priced correctly by the model by discounting cash-flows. Since we do not use options to calibrate the model, we view the algorithm as a way of generating a *curve of instantaneous forward rates* from the discrete dataset. In other words, we are primarily concerned with the modeling of "straight" debt instruments and not the study of the volatility of the forward rate curve. The curve is generated by the formula

$$\begin{aligned} f(T) &= -\frac{\partial}{\partial T}\log P(T)\\ &= -\frac{\partial}{\partial T}\log \mathrm{E}^{P}\left\{e^{-\int_0^T r_t\,dt}\right\}.\end{aligned}$$

where $f(T)$ and $P(T)$ represent the instantaneous forward rate and the discount factor (present value of a dollar) associated with the maturity date T. The instantaneous forward-rate curve allows us to price arbitrary fixed-income securities without optionality. Hedge-ratios for different instruments are derived from the sensitivities of the curve to input prices.

We consider a prior distribution for the short-term interest rate

$$\frac{dr_t}{r_t} = \sigma\, dZ_t + \mu_t^{(0)}\, dt\,, \tag{33}$$

where σ is constant and $\mu_t^{(0)}$ is given. For simplicity, we take $\mu_t^{(0)} \equiv 0$ under the prior, which, as we shall see, corresponds essentially to a prior belief of a flat forward-rate curve. [15]

Given the considerations of the previous sections, the family of candidate probability measures for the short rates has the form (33) where $\mu^{(0)}$ is replaced by an unknown drift μ_t.

The modified entropy functional (21) with $\eta = \frac{1}{2}m^2$ is

$$\begin{aligned} H_{mod}(P\,|\,P_0) &= \frac{1}{2\sigma^2}\,\mathrm{E}\left\{ \int_0^{T_{max}} e^{-\int_0^t r_s\,ds} \left(\mu_t - \mu_t^{(0)}\right)^2 dt \right\} \\ &= \frac{1}{2\,\sigma^2}\,\mathrm{E}\left\{ \int_0^{T_{max}} e^{-\int_0^t r_s\,ds}\, \mu_t^2\, dt \right\}. \end{aligned} \tag{34}$$

We calibrated this model to a data-set extracted from the US LIBOR market in late November 1997, consisting of FRAs and swap rates; cf. Table 1. The futures data corresponds to a series of 3-month Eurodollar contracts from January 1998 to December 2002. Forward-rates were computed from futures prices using an empirical convexity adjustment, which is displayed on the left of the futures price. Swap rates were computed from Treasury bond yields adding the corresponding credit spread, also displayed on the right of the yield. [16] Accordingly, the 3-month forward rate four months from today is computed as follows:

$$\begin{aligned} \text{forward rate} &= \text{futures-implied rate} - \text{conv. adjustment} \\ &= (100 - 94.20) - 0.12 \\ &= 5.68\ \% \end{aligned}$$

The 6-year swap rate was taken to be

$$\begin{aligned} \text{swap rate} &= \text{Treasury yield} + \text{spread} \\ &= 5.8150 + 0.3975 \\ &= 6.2125\ \% \end{aligned}$$

[15] Of course, we could have chosen any other drift for prior probability on short rates– this constitutes the "subjective" portion of the method. The significance of different priors will be clarified below.

[16] We shall not be concerned here about how convexity adjustments were generated or about the computation of the spread between swaps and Treasurys.

Table 1: Data for US LIBOR Market

ED futures / FRAs			Bonds / Swaps		
04m	94.20	0.0012	06y	5.8150	0.3975
10m	94.14	0.0023	07y	5.8236	0.4150
13m	94.08	0.0030	10y	5.8470	0.4475
16m	93.98	0.0044	12y	5.8683	0.4700
19m	93.98	0.0092	15y	5.9002	0.4800
22m	93.94	0.0131	20y	5.9535	0.4750
25m	93.91	0.0176	30y	6.0600	0.3750
28m	93.85	0.0234			
31m	93.87	0.0232			
34m	93.85	0.0371			
37m	93.83	0.0447			
40m	93.77	0.0522			
43m	93.79	0.0637			
46m	93.77	0.0730			
49m	93.75	0.0830			

In implementing the calibration algorithm for these instruments, we assumed that the discounted cash-flows of the FRAs per dollar notional are given by

$$\Gamma_f \;=\; e^{-\int_0^T r_t\,dt} \;-\; e^{-\int_0^{T+0.25} r_t\,dt}\left(1 \;+\; \frac{FRA \times 0.25}{100}\right)$$

where FRA is the 3-month forward rate (expressed in percentages) corresponding to the maturity T. The discounted cash-flows of a semi-annual vanilla interest swap with N cash-flow dates is

$$\Gamma_s \;=\; 1 \;-\; \sum_{n=1}^{N} e^{-\int_0^{0.5\,n} r_t\,dt}\left(\frac{SWAP \times 0.5}{100}\right) \;-\; e^{-\int_0^{0.5\,N} r_t\,dt} \;,$$

where $SWAP$ is the swap rate and where we assumed that the floating leg of the swap is valued at par.

In both cases (FRAs, swaps) we assumed that, under the risk-neutral probability, we have

$$\mathrm{E}^P\,\{\,\Gamma_i\,\} \;=\; 0 \qquad i \;=\; f,\, s\;.$$

These equations represent the constraints for calibration in this context. We have therefore 22 constraints: 15 for the FRAs and 7 for the swaps. The entropy-minimization was implemented by solving the partial differential equations (23), (24), (25) using a finite-difference scheme (trinomial lattice) and using L-BFGS to find the minimum of the augmented Lagrangian. We assumed a discretization of 12 periods per year.

Figure 1 shows the corresponding forward rate curve which derives from the data. We assumed a value of $\sigma = .10$ in this calculation. We noticed that the sensitivity of the curve to σ is negligible for $\sigma \leq 10\%$. The hedging properties of the model can be quantified by analyzing the sensitivities of the prices of par bonds with N years to maturity, for $N = 1, 2, 3 \dots 30$. These results are exhibited in the bar graphs diplayed hereafter. Each chart considers a par swap with a give maturity. The bars on the graph represent the sensitivity of the price of the instrument with respect to the prices of the input securities. Notice, in particular that the maturities that correspond to an input security consist of a single column. Intermediate maturities (not represented in the input instruments) give rise to multiple bars that decay as we move away from the corresponding maturity.

Finally, we point out that the volatility parameter σ in this model has an interesting interpretation. Heuristically speaking, the construction of the forward rate curve can be viewed as a problem in interpolation from a discrete set of data. Since the problem is ill-posed, various regularizations have been proposed at the level of forward-curve building, without having recourse to an underlying probability model. These regularizations typically penalize oscillations in the curve by means of penalization functions of the form

$$\int_0^{T_{max}} \eta(f(t),\, f'(t),\, f''(t),\, t)\; dt$$

that are typically minimized subject to the constraints and to a choice of function space for $f(t)$.

It is easy to see that, in the limit $\sigma \ll 1$, the minimum-entropy calibration algorithm is associated with a special choice of the above functional, namely,

$$\int_0^{T_{max}} e^{-\int_0^t f(s)\,ds} \left(\frac{f'(t)}{f(t)} - \mu^{(0)} \right)^2 dt\ . \tag{35}$$

This can be seen from the results of Section 5 and by letting σ formally tend to zero in the entropy functional

$$\mathrm{E}^P \left\{ \int_0^{T_{max}} m^2(t)\,dt \right\} = \frac{1}{\sigma^2}\, \mathrm{E}^P \left\{ \int_0^{T_{max}} \Big(\mu(t) - \mu^{(0)}(t) \Big)^2 dt \right\} .$$

This result corresponds mathematically to the relation between the “viscous” solution of the penalized problem associated with (35) and the stochastic control problem discussed in Section 5. Form a numerical point of view, we can therefore view the minimum-entropy algorithm as an “artificial viscosity” method for minimizing the functional (35) subject to the price constraints.

References

Avellaneda, M., Friedman, C., Holmes, R. and Samperi, D. (1997), Calibrating volatility surfaces via relative entropy minimization, *Applied Mathematical Finance*, March issue

Buchen, P. W. and Kelly, M.(1996), The maximum entropy distribution of an asset inferred from option prices, *Journal of Financial and Quantitative Analysis*, 31 (1).

Cover, J. and Thomas, J. A. (1991), *Elements of Information Theory*, New York: John Wiley & Sons

Gulko, L. (1995), The Entropy Theory of Option Pricing , Yale University Working Paper; (1996) The entropy theory of bond pricing, *ibid.*

Jackwerth, J. C. and Rubinstein, M. (1996), Recovering probability distributions from contemporaneous security prices, *Journal of Finance*, 69(3), 771-818

Jaynes, E.T., (1996), Probability theory: the logic of science, Unpublished manuscript, Washington University.

McLaughlin, D. W. (1984), *Inverse problems*, Providence RI: SIAM and AMS, Volume 14

Platen, E., Rebolledo, R. (1996), Principles for modelling financial markets, Advances in Applied Probabilitys 33(3) 601-613

Rockafellar, R.T. (1970), *Convex Analysis*, Princeton, NJ: Princeton University Press.

Samperi, D.(1997), *Inverse problems and entropy in Finance*, Ph.D. Thesis, New York University

Zhu, C., Boyd, R.H., Lu, P. and Nocedal, J. (1994),*L-BFGS-B: FORTRAN Subroutines for Large-Scale Bound-Constrained Optimization*, Northwestern University, Department of Electrical Engineering

Marco Avellaneda
Courant Institute
of Mathematical Sciences
New York University,
251 Mercer Street
New York, NY, 10012, USA