Minimum-Entropy Calibration of Asset-Pricing Models *

Marco Avellaneda †

Abstract

We present an algorithm for calibrating asset-pricing models to benchmark prices. The algorithm computes the probability that minimizes the relative entropy with respect to a prior and satisfies a finite number of moment constraints. These constraints arise from fitting the model to the prices of benchmark instruments. Generically, there exists a unique solution which is stable, *i.e.* it depends smoothly on the input prices. We study the sensitivities of the values of contingent claims with respect to variations in the benchmark prices in detail. We find that the sensitivities can be interpreted as regression coefficients of the payoffs of contingent claims on the set of payoffs of the benchmark instruments, under the resik-neutral measure. 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 swap/FRA data and to study the corresponding sensitivities of fixed-income securities to variations in input prices.

*The author thanks Craig Friedman, Richard Holmes, Dominick Samperi, Joseph Langsam, Les Gulko and an anonymous referee for valuable discussions and suggestions. This paper was presented at several seminars and conferences in early 1998 including the New York IAFE/CIFRE Conference, the Columbia-JAFFE Conference on Mathematical Finance, the Eugene Fabes Memorial Conference in the Politecnico di Milano, the Venice-2 Symposium on Applied Mathematics, the Journee Annuelle de la Societe Mathematique de France in Paris and the Quantitative Finance Conference, City University, Hong Kong. The author thanks the organizers and participants of these events and for their enlightening comments.

†Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY, 10012

1 Introduction

Despite its practical importance, model calibration has received little attention in Mathematical Finance theory. Calibrating an asset-pricing model means specifying a probability distribution for the underlying variables in such a way that the model reproduces, by taking discounted expectations, the current market prices of a set of benchmark instruments. 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 "local volatility" models used in option portfolio management.[2]

In many cases of interest, calibrating a model is equivalent to solving a classical problem in statistics: the determination of a probability distribution from a finite set of moments. Here, the "moments" correspond to the discounted expectations of the cash-flows of the benchmark instruments. It is well-known, however, that this problem is ill-posed: sometimes there can be multiple solutions and 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 benchmark 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 a algorithm for selecting a risk-neutral probability that minimizes the *relative entropy*, or *Kullback-Leibler entropy*, with respect to a given prior distribution. 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 *a-priori* information (*e.g* of econometric nature)that the modeler may have. This information is modeled by the prior probability. The entropy-minimization algorithm provides a way of reconciling the prior information (econometric, historic) with the information contained in current prices of benchmarks.[4]

Buchen and Kelly(1996) and Gulko(1995, 1996, 1998) 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

[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".

[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.

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

the calibration of volatility surfaces of multiperiod option pricing models. In the present paper, we consider a general intertemporal asset-pricing model and use the method of Lagrange multipliers to model price constraints, following Buchen and Kelly(1996) and Avellaneda *et al*(1997). The calibration algorithm consists in finding the minimum of the augmented Lagrangian. We also study the sensitivities of the prices of contingent claims with respect to perturbations in the input prices. This sensitivity analysis uses the the matrix of second derivatives (Hessian) of the Lagrangian computed at the critical point.

One of the main results of this study is a characterization of the price-sensitivities obtained when using the min-entropy method. Consider a "generic" contingent claim which is not one of the benchmark instruments. Let Π denote the model price of the contingent claim and let h represent its discounted payoff. Further, let us denote by $G_i,\ i = 1, 2, ..., N$ the discounted cash-flows of the benchmark instruments, and by $C_1, \ldots C_N$ their prices. We will establish in that

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

where

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

and **Cov** represents the covariance operator under the risk-neutral (calibrated) measure.

This result has a clear financial interpretation. It is well-known that the right-hand side of the first equation corresponds to the value of th 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. Thus, the *deltas* (price-sensitivities) and the *betas* (the regression coefficients of the cash-flow of the contingent claim on the set of cashflows of the benchmark instruments) are identical! This property of the minimum-entropy algorithm suggests that the model produces hedging strategies that are consistent with a prior statistical distribution of the underlying factors. [5]

The paper is organized as follows: In Section 2, we consider a one-period model. Under mild assumptions, we show that if there exists a probability with finite relative entropy, then the min-entropy calibration problem has a unique solution. We establish

[5]As opposed to, say, a probability distribution that is calibrated to the prices of benchmark instruments but inconsistent with stylized facts about the market.

also that the price-sensitivities of contingent claims depend smoothly on the input prices.

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[6] 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. Specifically, minimizing relative entropy is equivalent to minimizing the L_2 norm of the risk-premia $m_i(t)$, *i.e.*

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

where T_{max} is the time-horizon and ν is the number of factors. This suggests that there are many generalizations of the "pure" entropy algorithm that can be made by changing the form of the cost function.

In practice, it is computationally advantageous to consider the cost function

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

This reduces the dimensionality of the computation, while preserving some of the essential features of the algorithm. Other modifications may involve replacing the quadratic form $\sum_i m_i(t)^2$ by a general convex function of the risk-premia.[7] We note, however, that the aforementioned "delta=beta" property of the hedge-ratios of the minimum-entropy algorithm no longer holds for the generalized algorithms.[8]

In Section 6, we apply the algorithm to construct forward rate curves from US LIBOR data (FRAs and swap rates). We pay particular attention to (a) the smoothness of the forward rate curve and (b) the price-sensitivities with respect to input swap rates. The analysis of the sensitivities resulting from different methods for generating forward-rate curves remains an interesting issue among practitioners. At stake is the quality of the hedges generated by different procedures. Hedges tend to be model-dependent and therefore a certain amount of risk is taken when choosing

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

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

[8]Nevertheless, since the algorithms are structurally similar, we expect that the latter will produce hedge-ratios that are, in some sense, close to the "risk-neutral betas". This is certainly to be expected if the calibrated probability measure is near the prior in the Kullback distance.

different forward rate curves. The issue is whether smooth curves, which give rise to "non-local" hedges[9], are preferable to discontinuous forward rate curves, such as the ones obtained by the bootstrapping method, or to the curves obtained by spline interpolation.

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

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

where $G_1(X)$, $...G_N(X)$ are given functions and C_1, $...C_N$ are given numbers. [10] Here, 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 (benchmarks).

Buchen and Kelly(1996) 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 $f_0(X)$. [11]

It is well-known (Cover and Thomas(1991)) 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

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

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

[10]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.

[11]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(1991)).

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 (4) into (3), 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.

A crucial feature of this algorithm is that the objective function is convex in λ. In fact, we have

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

Since covariance matrices are non-negative definite, $log(Z(\lambda)) \; - \; \lambda \cdot C$ is convex. It also follows that $log(Z(\lambda))$ is strictly convex if the N payoff functions are linearly independent. [12]

The stability of the solution, i.e. the continuous dependence of f_λ on input prices, follows from convex duality. Let λ^* be the value of the Lagrange multipliers that minimizes the objective function $log\left[Z(\lambda)\right] \; - \; \lambda\, C$. To assess the sensitivity of the

[12] 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.

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

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

Hence,

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

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

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

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:

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

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

We assume that there are N benchmark 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

$$C_i \;=\; \mathbf{E}^P \left\{ \sum_{k=1}^{n_i} e^{-\int_0^{T_{ik}} r(s)\, ds} G_{ik}(\mathbf{X}(T_{ik})) \right\} \;, \qquad i = \;1,\; 2,\; ...\; N \;, \tag{8}$$

where $\{T_{ik}\}_{k=1}^{n_i}$ are the cash-flow dates of the i^{th} security and $\{G_{ik}(\mathbf{X})\}_{k=1}^{n_i}$ represent the corresponding cash-flows. We assume that the latter are bounded, continuous functions of $\mathbf{X}$. The process $r(s) \;=\; r(\mathbf{X}(s), s)$ 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

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

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} . [13]

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

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

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))

$$Z(\lambda) \; = \; \mathbf{E}^{P_0} \left\{ \exp \left(\sum_{i=1}^{N} \lambda_i \sum_{j=1}^{n_i} e^{-\int_0^{T_{ij}} r(s) ds} G_{ij}(\mathbf{X}(T_{ij})) \right) \right\} . \tag{10}$$

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:

$$\frac{dP_\lambda}{dP_0} \; = \; \frac{1}{Z(\lambda)} \cdot \exp \left(\sum_{i=1}^{N} \lambda_i \sum_{j=1}^{n_i} e^{-\int_0^{T_{ij}} r(s) ds} G_{ij}(\mathbf{X}(T_{ij})) \right) . \tag{11}$$

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

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

$$C_i \ = \ \frac{Z_{\lambda_i}}{Z} \ = \ \mathbf{E}^{P_\lambda} \left\{ \sum_{j=1}^{n_i} e^{-\int\limits_0^{T_{ij}} r(s)ds} G_{ij}(\mathbf{X}(T_{ij})) \right\} \qquad 1 \le i \le N \ .$$

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

Define the discounted cash-flows

$$\Gamma_i \ = \ \sum_{j=1}^{n_i} e^{-\int\limits_0^{T_{ij}} r(s)ds} G_{ij}(\mathbf{X}(T_{ij})) \ , \quad 1 \ \le \ i \ \le \ N \ .$$

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

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

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

$$\frac{\partial}{\partial \lambda_j} \mathbf{E}^P \left\{ e^{-\int\limits_0^T r_s \, ds} h(X_T) \right\} \ = \ \mathbf{Cov}^P \left[\Gamma_j, \ e^{-\int\limits_0^T r_s \, ds} h(X_T) \right] \ .$$

Thus, the results obtained in Section 2 apply to the general intertemporal asset pricing model (7) and we have

Proposition 2. *(a) Minimizing the relative entropy for the model (7) subject to the price constraints (8) is equivalent to searching for a probability measure belonging to the N-parameter family (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 $\mathbf{L}^{(0)}$ represent the infinitesimal generator of the semi-group corresponding to the prior P_0 *i.e.*,[14]

$$\mathbf{L}^{(0)}\,\phi \;=\; \frac{1}{2}\sum_{ij=1}^{\nu} a_{ij}\,\phi_{X_i\,X_j} \;+\; \sum_{i=1}^{\nu}\mu_i^{(0)}\,\phi_{X_i} \tag{12}$$

where

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

It is useful to introduce the auxiliary state-variable $Y_t \;=\; e^{-\int_0^t r(s)\,ds}$. Notice that $(\mathbf{X}_t,\,Y_t)$ is a $\nu+1$-dimensional Markov process with infinitesimal generator

$$\mathbf{L}^{(0)} \;-\; r\,Y\,\frac{\partial}{\partial Y}\;.$$ [15]

It follows from (10) and standard diffusion theory (Feynman-Kac formalism) that the normalization factor $Z(\lambda)$ is given by

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

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

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

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

$$U(\mathbf{X},\,Y,\,T_{ij}-0\,;\lambda) \;=\; U(\mathbf{X},\,Y,\,T_{ij}+0\,;\lambda)\cdot\exp\left(\sum_{i=1}^{N}\lambda_i\,G_{ij}(\mathbf{X})\,Y\right)\;. \tag{15}$$

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

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

[15] The reason for introducing the "augmented" state-process $(\mathbf{X}_t,\,Y_t)$ is that relative entropy is a quantity independent of the discount rate. As we shall see later, it is possible to simplify the PDE by modifying the notion of entropy.

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

where W satisfies the PDE

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

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

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

$$\sum_{j=1}^{n_l} G_{lj}(\mathbf{X})\,Y\,\delta(t-T_{lj})\;. \tag{17}$$

Numerically, the algorithm consists in finding the minimum of $W(\mathbf{X}(0),\,1,\,0;\,\lambda) - \lambda{\cdot}C$ with respect to λ. This is done iteratively, by solving the the PDEs (16) and (17) for the function and its gradient, given by $V^{(l)}(\mathbf{X}(0),\,1,\,0;\,\lambda) \;-\; C_l,\;\; l=1,\,...,\,N$. At the critical point, the first-order optimality condition gives $V^{(l)} \;-\; C_l \;=\; 0$, as desired.

From the drift term of equation (17), 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(1996)):

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 $0 < t < T_{max} = \max_{ik} T_{ik}$) is given by

$$H(P|P_0) \ = \ \frac{1}{2}\, \mathbf{E}^P \left\{ \int_0^{T_{max}} \sum_{j=1}^{\nu} m_j(t)^2\, dt \right\} \ . \tag{19}$$

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

$$-\mathbf{E}^P \left[\int_0^{T_{max}} \sum_{j=1}^{\nu} m_j^2(t)\, dt \right] + \sum_{i=1}^{N} \lambda_i\, \mathbf{E}^P \left[\sum_{j=1}^{n_i} e^{-\int_0^{T_{ij}} r(s)ds}\, G_{ij}(\mathbf{X}(T_{ij})) \ - \ C_i \right] \tag{20}$$

The advantage of the stochastic control formulation is that it can be generalized considerably. In fact, we can replace the function $\sum_j m_j^2(t)$ 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

$$H_{mod}(P|P_0) \; = \; \frac{1}{2}\,\mathbf{E}\left\{ \int\limits_0^{T_{max}} e^{-\int\limits_0^t r(s)\,ds} \; \eta(m(t)) \;\; dt \right\} \,, \tag{21}$$

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

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

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

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 $\eta(t,\, m) \; = \; \frac{1}{2} \sum\limits_j m_j^2$ we have $\eta = \eta^*$. The corresponding Bellman equation is

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

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

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

$$W\,(\mathbf{X}(0),\, 0\,;\; \lambda_1,\; ...\; \lambda_N) - \sum_{i=1}^{N} \lambda_i \, C_i \;,$$

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:

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

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

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

$$V^{(l)}\,(\mathbf{X}(0),\,0\,;\ \lambda_1,\ \ldots\ \lambda_N)\ -\ \lambda_i\,C_l\ =\ 0\ ,\qquad 1\ \le\ l\ \le\ N\ .$$

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

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

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

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

In particular, we have

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

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

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

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

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

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

In particular, this shows that if the instruments are not linearly dependent with P_0-probability 1, the Hessian matrix is positive definite.[16] [17] 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

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

and

$$\Pi_t^{(l)} \;+\; \mathbf{L}^{(0)}\,\Pi^{(l)} \;+\; \sum_{ij} a_{ij}\, W_{X_i}\, \Pi_{X_j}^{(l)} \;+\; \sum_{ij} a_{ij}\, \Pi_{X_i}\, V_{X_j}^{(l)} \;-\; r\,\Pi^{(l)} \;=0\;. \quad (30)$$

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

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

[17] 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.

$$
\begin{aligned}
\Pi^{(l)} &= \mathbf{E}^{P_\lambda}\left\{ \int_0^{T_{max}} e^{-\int_0^t r(s)\,ds} \sum_{ij=1}^{\nu} a_{ij}\, V_{X_i}^{(l)}\, \Pi_{X_j}\ dt \right\} \\
&= \mathbf{Cov}^{P_\lambda}\left[e^{-\int_0^T r(s)\,ds} h(\mathbf{X}_T),\quad \Gamma^{(l)} \right] + \\
\mathbf{E}^{P_\lambda}&\left\{ \int_0^{T_{max}} \left(e^{-\int_0^t r(s)\,ds} - e^{-2\int_0^t r(s)\,ds} \right) \sum_{ij=1}^{\nu} a_{ij}\, V_{X_i}^{(l)}\, \Pi_{X_j}\ dt \right\}. \qquad (31)
\end{aligned}
$$

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

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

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 benchmark instruments in the US dollar swap market.

We consider a set of benchmark instruments consisting of forward-rate agreements (FRAs) and swaps with standard maturities. Using the algorithm, we compute a probability measure for the process driving the short-term rate such that all the benchmark instruments are priced correctly by discounting cash-flows. Since we do not use options, we view the algorithm in this context as an alternative way of generating a curve of instantaneous forward rates.

The curve is generated by the formula

$$f(T) = -\frac{\partial}{\partial T} \log P(T)$$

$$= -\frac{\partial}{\partial T} \log \mathbf{E}^P \left\{ e^{-\int_0^T r_t\, dt} \right\} .$$

where $f(T)$ and $P(T)$ represent the instantaneous forward rate and the discount factor (present value of \$1) associated with the maturity date T. [18] The forward-rate curve allows us to price arbitrary fixed-income securities without optionality. Hedge-ratios for different instruments are derived from the sensitivities to benchmark 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 $\mu_t^{(0)} = 0$ under the prior, which, as we shall see, corresponds to a flat forward-rate curve for $\sigma \ll 1$. [19]

From the considerations of the previous sections, the candidate probability measures for the short rate process have the form (33), with $\mu^{(0)}$ replaced by an unknown drift μ_t. The volatility σ is fixed and plays the role of an adjustable parameter.

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

$$H_{mod}(P\,|\,P_0) = \frac{1}{2\sigma^2} \mathbf{E} \left\{ \int_0^{T_{max}} e^{-\int_0^t r_s\, ds} \left(\mu_t - \mu_t^{(0)} \right)^2 dt \right\}$$

$$= \frac{1}{2\,\sigma^2} \mathbf{E} \left\{ \int_0^{T_{max}} e^{-\int_0^t r_s\, ds} \mu_t^2\, dt \right\} . \tag{34}$$

We calibrated this model to benchmark FRAs and swap rates the US dollar LIBOR market on a particular date in late November 1997; cf. Table 1. The futures prices correspond to 3-month Eurodollar futures contracts from January 1998 to December 2002. Forward-rates were computed from futures prices using an empirical convexity

[18] Throghout this section, the letter f denotes the instantaneous forward-rate curve. This should not be confused with the notation of §2, where f represented a probability density.

[19] Of course, we could have chosen any other drift for prior probability – this constitutes the "subjective", or "econometric" portion of the method. For example, we could start with a mean-reverting short-rate model (Vasicek (1977)) in which the volatility, rate of mean-reversion and asymptote for the forward rate are determined from historical data. The significance of using different priors $\mu^{(0)}$ is clarified at the end of the section.

adjustment, displayed immediately on the right 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. [20] 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

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

[20] 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

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

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. The tolerance of the BFGS algorithm was such that the prices of the benchmark instruments were matched with an error of a fraction of a basis point ($0.0001) on a notional of $100.

Figure 1 shows the corresponding forward rate curve which derives from the data. We assumed a value of approximately $\sigma = .10$ in this calculation. We noticed that the sensitivity to the value of σ is small at these levels. One of the effects of having a large σ versus a small one is to accentuate the smoothness of the curve.

We analyze the the hedge-ratios produced by this model for hedging straight by computing the sensitivities to variations in the benchmarks of the prices of par bonds with N years to maturity, for $N = 1, 2, 3 \ldots 30$. These results are exhibited in the bar charts displayed hereafter. Each chart considers a par swap with a given maturity. The bars orepresent the sensitivity of the price of the instrument with respect to the prices of the 22 benchmarks. Notice, in particular that the maturities that coincide with a benchmark consist of a single column (100% sensitivity). Intermediate maturities (not represented in the list of input instruments) give rise to charts with multiple bars which decrease in magnitude as we move away from the maturity that we are trying to hedge.

Finally, let us show that this model has some properties that connect it with more traditional ways of building the forward-rate curve. Heuristically, the construction of the forward rate curve can be viewed as an interpolation problem where we start from a discrete set of prices and derive a continuum of security prices. Since the problem is ill-posed. These approaches usually introduce penalty functions of the form

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

that are minimized subject to the constraints or, as a regularization term for least-squares fitting. The most common penalization functions are $\eta = (f')^2$ and $\eta = (f'')^2$, which correspond, respectively to quadratic and cubic "tension splines".

The relation between the min-entropy model and tension splines models is seen for small values of the volatility parameter. It is easy to see that, for $\sigma \ll 1$, the solution of the minimum-entropy calibration algorithm converges to the solution of a deterministic optimization problem for the functional

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

where $\mu^{(0)} = (\log f_0))'$ is logarithmic derivative of a "prior" forward-rate curve. [21] This functional can therefore be regarded as a particular choice of "tension splines".

[21] In particular, the prior drift $\mu^{(0)} = 0$ corresponds a flat forward-rate curve.

The above result is obtained by letting σ formally tend to zero in the rescaled entropy functional

$$\sigma^2\ \mathbf{E}^P \left\{ \int_0^{T_{max}} m^2(t)\, dt \right\} \ = \ \mathbf{E}^P \left\{ \int_0^{T_{max}} \Big(\mu(t) \ - \ \mu^{(0)}(t) \Big)^2 \ dt \right\} \ . \qquad (36)$$

Mathematically, this reflects the correspondence between the "viscosity solution" of the deterministic optimization problem associated with (35) and the stochastic control problem for the functional (36). [22]

Therefore, we can interpret the the minimum-entropy algorithm in the present context as an "articifial viscosity" computational method for fitting a forward rate curve to benchmark prices by minimizing the functional (35). Notice that we can modify the functional form of the stochastic differential equation for $r(t)$ (33) to obtain different regularization functions depending on f and f' in the limit $\sigma \ll 1$. This leads usto the question of comparing the min-entropy method with other approaches for constructing forward-rate curves, an issue left to a future publication.

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* ; (1998), *The Entropy Pricing Theory*, Ph. D. Thesis, Yale School of Management, Yale University

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

[22] A rigorous justification of the asymptotics leading to (35) can be made from these considerations.

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

Vasicek, O. (1977), "An equilibrium characterization of term-structure", *Journal of Financial Economics*, 5, 177-188

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

7 Appendix: minimization of relative entropy under moment constraints

For the sake of completeness, we include a brief proof of equation (4), Section 2. The issue is to characterize the solution of the variational problem

$$\sup_f \left[-\int log\left(\frac{f}{f_0}\right) f\,dX \quad + \sum_i \lambda_i \left(\int G_i\, f\, dX \; - \; C_i \right) \right] ,$$

where f varies over all probability densities.

We incorporate the constraint

$$\int f\; dX \;=\; 1$$

as an additional Lagrange multiplier, i.e. we consider the augmented Lagrangian

$$-\int log\left(\frac{f}{f_0}\right) f\,dX \quad + \sum_i \lambda_i \left(\int G_i\, f\, dX \; - \; C_i \right) \; + \; \mu \; \left(\int f\; dX \; - \; 1 \right) \; .$$

Taking the first variation with respect to f, we obtain the first-order condition

$$-\int log\left(\frac{f}{f_0}\right) \delta f\,dX \quad - \int \delta f\; dX \; + \sum_i \lambda_i \int G_i\, \delta f\, dX \; + \; \mu \int \delta f\; dX \quad = \; 0$$

for all variations δf. Since this has to hold for all variations δf, the critical f must satisfy

$$-log\left(\frac{f}{f_0}\right) - 1 \; + \; \sum_i \lambda_i\, G_i \; + \; \mu \; = \; 0,$$

or,

$$f(X) \;=\; e^{\mu \,-\, 1} \; e^{\sum\limits_i \lambda_i \; G_i(X)} \; f_0(X)\;.$$

The prefactor is determined by the requirement that the integral of f be equal to 1. We conclude that the critical point of the above functional is precisely

$$f_\lambda \;=\; \frac{e^{\sum\limits_i \lambda_i \; G_i} \; f_0}{\int e^{\sum\limits_i \lambda_i \; G_i} \; f_0 \; dX}\;.$$

This is what we wanted to show.