

PROBABILITY DISTRIBUTIONS OF ASSETS INFERRED FROM OPTION PRICES VIA THE PRINCIPLE OF MAXIMUM ENTROPY*

J. BORWEIN[†], R. CHOKSI[‡], AND P. MARÉCHAL[§]

Abstract. This article revisits the maximum entropy algorithm in the context of recovering the probability distribution of an asset from the prices of finitely many associated European call options via partially finite convex programming. We are able to provide an effective characterization of the constraint qualification under which the problem reduces to optimizing an explicit function in finitely many variables. We also prove that the value (or objective) function is lower semicontinuous on its domain. Reference is given to a website which exploits these ideas for the efficient computation of the maximum entropy solution (MES).

Key words. European options, maximum entropy, semifinite programming, Lagrangian duality, convex conjugate

AMS subject classifications. 90C25, 49N15, 91B28

DOI. 10.1137/S1052623401400324

1. Introduction. Entropy optimization, used for recovering a probability distribution from information on a few of its moments, is well established and ubiquitous throughout the sciences [14]. Recently (cf. Buchen and Kelly [9] and Avellaneda et al. [1], [2]), this idea has been explored in the context of financial derivatives. In this *risk-neutral* model, one wishes to infer the probability distribution for the price of an asset at some future date T from the prices of European call options based upon the asset with expiration at T .

A classical approach to the application of entropy optimization has been to use the theory of Lagrange multipliers. While this formal approach does yield correct and useful results, it does not provide for a complete analysis. The purpose of this article is to analyze the option-maximum entropy problem within the framework of partially finite programming and demonstrate the extra insight and power that this approach provides. In doing so, we not only legitimize the formal calculations with Lagrange multipliers but also provide a more detailed analysis of the maximum entropy solution and the notion of admissible data. We also specifically exploit the unique structure of the piecewise linear constraints to reduce the problem to maximization of an *explicit function of finitely many variables*; hence greatly simplifying the computation of the maximum entropy solution.

The option-maximum entropy problem. Let I be an the interval of the form $[0, K)$ with either some fixed $K > 0$ or $K = +\infty$. For $0 = k_1 < k_2 < \dots < k_m$, and

*Received by the editors December 28, 2001; accepted for publication (in revised form) January 15, 2003; published electronically October 14, 2003.

<http://www.siam.org/journals/siopt/14-2/40032.html>

[†]Department of Mathematics and Centre for Experimental and Constructive Mathematics, Simon Fraser University, Burnaby, Canada (jborwein@cecm.sfu.ca).

[‡]Department of Mathematics, Simon Fraser University, Burnaby, Canada (choksi@math.sfu.ca).

[§]Département de Sciences Mathématiques, Université Montpellier II, Montpellier, France (marechal@darboux.math.univ-montp2.fr).

$\mathbf{d} \in \mathbb{R}^m$,

$$(\mathcal{P}) \quad \left\{ \begin{array}{l} \text{minimize} \quad \mathcal{I}_h(p) := \int_I h(p(x)) \, dx \\ \text{s.t.} \quad 1 = \int_I p(x) \, dx, \\ \quad \quad d_j = \int_I c_j(x)p(x) \, dx. \end{array} \right.$$

Here, $p(x)$ denotes the *probability density function* for the *price* x of an *asset* at a set *future time* T , and d_j represents the price of a *European call option* based on the underlying asset with *strike price* k_j and *expiration date* T . The interval I denotes the set of feasible prices for the asset at time T which may or may not be a priori constrained. The function $c_j(x)$ represent the payoffs of the j th option as a function of the asset price x at time T . Thus

$$(1) \quad c_j(x) = (x - k_j)^+ = \max\{0, x - k_j\}.$$

Finally the convex function $h : \mathbb{R} \rightarrow \mathbb{R}$ represents the entropy functional, the most common of which being the *Boltzmann-Shannon entropy*

$$(2) \quad h(t) := \begin{cases} t \log t - t & \text{if } t > 0, \\ 0 & \text{if } t = 0, \\ +\infty & \text{if } t < 0. \end{cases}$$

Note that traditionally, the entropy is taken to be $-h$, and hence maximum entropy entails solving for the minimum in (\mathcal{P}) . We refer to the minimizer associated with (\mathcal{P}) as the maximum entropy solution, or simply the MES.

The particular choice of the Boltzmann-Shannon entropy yields a simple case of the *minimum cross entropy problem* using the *Kullback-Leibler* entropy. Here the idea is that given additionally a prior guess $q(x)$ for the asset price distribution at T (which one might infer from the market), one seeks to find the least prejudiced posterior density $p(x)$ consistent with the constraints which is *closest* to or least *deviant* from $q(x)$ in the following sense (see Cover and Thomas [10] for details): find a constraint satisfying $p(x)$ which minimizes

$$\int_I p(x) \log \left(\frac{p(x)}{q(x)} \right) \, dx.$$

Our problem (\mathcal{P}) is the simple case of the above where no prior is available and hence $q(x)$ is close to a uniform distribution and may be taken to be a constant. Of course to be precise, it will be uniform if $p(x) = 0$ for all x sufficiently large (cf. [9]). For simplicity we first carry out our analysis for the Boltzmann-Shannon entropy (i.e., uniform prior). In section 7, we briefly comment on the necessary modifications and drawbacks in the more realistic situation of including a nontrivial prior.

The constraints in (\mathcal{P}) may appear to be missing something. Indeed, they should read

$$d_j = DC(T) \int_I c_j(x)p(x) \, dx,$$

where $DC(T)$ represents the *riskless discount factor* up to time T . For example, one could take

$$DC(T) = e^{-rT},$$

where r is the *risk-free constant interest rate*. Without loss of generality we set $DC(T) = 1$ throughout this paper. Finally, we emphasize that this model, in which the option prices are simply the expected values of a discounted pay-off function, assumes *risk-neutrality*. See [13] (also [9] and the references therein) for further information on *risk-neutral* pricing and *arbitrage-free* models.

Convex programming approach. In this article, we reexamine problem (\mathcal{P}) within the general framework of convex duality and partially finite convex programming. Why this approach? To begin with, it legitimizes the calculations done in [9] and [1] which are based upon Lagrange multipliers. This standard approach is based on relaxing the hard constraints via Lagrange multipliers, reducing the problem to

$$(3) \quad \sup_{\boldsymbol{\lambda} \in \mathbb{R}^{m+1}} \inf_p L(\boldsymbol{\lambda}, p),$$

where

$$(4) \quad L(\boldsymbol{\lambda}, p) := \mathcal{I}_h(p) + \lambda_0 \left(1 - \int_0^\infty p \, dx \right) + \sum_{i=1}^m \lambda_i \left(d_i - \int_0^\infty c_i p \, dx \right).$$

The minimization over p is carried out via the first variation of $L(\boldsymbol{\lambda}, p)$ with respect to p ; i.e., one “differentiates” the Lagrangian with respect to density functions p . There is a problem with this type of calculation. Indeed, the Lagrangian has support on the set $\{p \in L^1(I) \mid p \geq 0 \text{ a.e.}\}$. The complement of this set is dense in L^1 and, moreover, any reasonable definition of the Boltzmann–Shannon entropy gives a value of $+\infty$ on any function in the complement. Thus, not only is the Lagrangian nowhere differentiable, it is indeed nowhere continuous. The approach via conjugation-duality is in part to circumvent this differentiation. Moreover, with other entropies, there can be additional complications to these formal calculations resulting from a lack of weak compactness. See [8] for a fuller discussion. We emphasize, however, that the benefits of our approach are far from confined to the legitimization of the now fairly ubiquitous if flawed formal analysis with “Lagrange multipliers.” Such benefits include the following:

- We transform the maximum entropy problem into a closed-form finite-dimensional maximization problem. That is, under certain explicit conditions on the data, finding the MES is equivalent to maximizing an explicit dual function (cf. (9), (12), and (16)) of finitely many real variables. The simple fact that the dual function can be written explicitly with no integrals is an advantage of using a uniform prior.
- Our approach greatly simplifies the numerical computation¹ of the MES where many of the previous numerical calculations (cf. [9]) involved in computing the optimal $\boldsymbol{\lambda}$ can now be done symbolically.
- We give a detailed analysis of the constraint qualification (CQ) and a full investigation of when the MES exists, and when the maximization with respect to $\boldsymbol{\lambda}$ in the dual (cf. (9), (12)) does indeed yield the solution. These results are pertinent when analyzing the dependence of the MES on the data \mathbf{d} .

¹An interface has been set up at <http://www.cecm.sfu.ca/projects/MomEnt+/moment.html> which computes the MES for a variety of moment constraints, including the ones discussed in the present paper. One can test our algorithm by first pricing the list of options using, for example, a log-normal distribution, and then comparing the distribution with the computed MES based only on the option prices. In this way, one finds that the accuracy of recovering a known distribution with eight options is quite high even with a uniform prior.

- Our general approach applies to any convex entropy, not just to the standard Boltzmann–Shannon entropy used in [9], [1]. It is also amenable to natural extensions such as relaxations of the constraints, for example, requiring the moments to lie in some small finite interval.
- Partially finite duality and attainment results are usually confined to primal function spaces defined over bounded domains. The problem provides an interesting and simple example whereby a partially finite duality and attainment theorem can be proved in the case where the primal functions are defined over an infinite domain (Theorem 2). We know of no general result which would capture this.

In section 3 we prove two duality results: one for the case of a finite interval I and the other for $I = [0, \infty)$. The first (Corollary 1) is a direct consequence of a well-known duality result (Theorem 1). The latter (Theorem 2) is proved directly by exploiting the monotonicity of the constraints c_i . In either case, the MES exists if d satisfies the CQ. Conversely, for the MES to exist in its exponential form (cf. (10), and (13)), this CQ must hold. The CQ amounts to the data \mathbf{d} lying in the relative interior of the *feasible set*, i.e., the set of vectors $(y_0, y_1, \dots, y_m) \in \mathbb{R}^{m+1}$ such that

$$y_i = \int_I c_i(x)p(x) dx \quad \text{for } i = 0, \dots, m$$

for some distribution p with finite entropy. In section 4 we show that this condition is equivalent to the data \mathbf{d} lying in some open polyhedral set which we characterize explicitly (cf. Proposition 2). It is important to note that the feasible set is not relatively open, and hence there can exist boundary points which are feasible even though the CQ fails. In such cases, the analysis via the Lagrange multipliers λ_i will fail. Indeed, as the data approaches such a boundary point, some components of the associated $\boldsymbol{\lambda}$ will become infinite.

We provide a simple—though perhaps artificial from a finance point of view—example to illustrate these points. We use only two constraints for simplicity (similar examples exist with many options) and assume the first option has strike price zero. That is, we consider strike prices $k_1 = 0, k_2$ with associated option prices d_1 and d_2 (with $d_2 \geq 1/2$). This data satisfies the CQ if and only if

$$0 < d_1 - d_2 < k_2.$$

The boundary point where $d_1 - d_2 = k_2$ is of particular interest. Clearly, this data is feasible; for example, consider

$$p = \chi_{[k_2+d_2-\frac{1}{2}, k_2+d_2+\frac{1}{2}]}$$

Moreover one can readily show (see (18)) that any probability distribution satisfying the associated constraints must vanish on the interval $[0, k_2]$. Hence, no MES solution of the exponential form (i.e., (10)) can exist. Indeed, as data satisfying the CQ tends to this boundary point, the associated $\boldsymbol{\lambda}$ must blow up. This simple example illustrates that an infimum associated with problem (\mathcal{P}) might still be finite but not attainable. In section 6, we explore this matter further by studying the *value (or objective) function* and whether or not there exists a *duality gap*.

2. Preliminaries. We first reformulate problem (\mathcal{P}) . Let $I = [0, K)$ with either $K > 0$ fixed or $K = +\infty$. For $m \geq 1$, we assume that $0 = k_1 < \dots < k_m < K$, and

$\mathbf{d} = (d_0, d_1, \dots, d_m) \in \mathbb{R}^{m+1}$ with $d_0 = 1$. Consider

$$(P) \quad \inf \{ \mathcal{I}_h(p) + \delta(\mathbb{A}p - \mathbf{d} \mid \mathbf{0}) \}, \quad \text{where } \mathcal{I}_h(p) := \int_I h(p(x)) dx,$$

δ is the indicator function defined for the set $\{\mathbf{0}\}$, i.e., for $\mathbf{y} \in \mathbb{R}^{m+1}$,

$$\delta(\mathbf{y} \mid \mathbf{0}) := \begin{cases} 0 & \text{if } \mathbf{y} = \mathbf{0}, \\ \infty & \text{otherwise,} \end{cases}$$

and \mathbb{A} is the linear operator defined by

$$\mathbb{A}p := \int_I \mathbf{c}(x)p(x) dx \in \mathbb{R}^{m+1},$$

with $\mathbf{c}(x) = (c_0(x), c_1(x), \dots, c_m(x))$, $c_0(x) \equiv 1$, $c_j(x) = (x - k_j)^+ = \max\{0, x - k_j\}$. Finally, h always denotes the Boltzmann–Shannon entropy defined by (2). The space we will work in for admissible p is $L^1(I)$. We will separate the cases of bounded I and $I = [0, \infty)$. For the latter case, \mathbb{A} may be infinite on some $p \in L^1([0, \infty))$, and hence for the problem at hand \mathbb{A} is not a well-defined linear operator on $L^1([0, \infty))$ as it would be on, say, $L^1([0, M])$ for some fixed $M > 0$. One notes that even though the operator \mathbb{A} is densely defined on $L^1([0, \infty))$, it is not closed. Hence this case requires a different approach. For the case of bounded I , we will directly apply partially finite convex programming (Theorem 1) to establish the duality relation under a CQ. A similar duality relation (Theorem 2), under the same CQ, holds true for the infinite domain $I = [0, \infty)$ and will be proved directly, bypassing the Fenchel duality of Theorem 1.

For omitted definitions and elementary facts from convex analysis in \mathbb{R}^n we refer the reader to [17]. Let V and V^* be vector spaces equipped with $\langle \cdot, \cdot \rangle$, a bilinear product on $V \times V^*$. The *convex (Fenchel) conjugate* of a convex function f on V with respect to $\langle \cdot, \cdot \rangle$ is the function f^* defined on V^* by

$$f^*(\xi) := \sup \{ \langle x, \xi \rangle - f(x) \mid x \in V \}.$$

We consider the functional on $L^1(I)$ (for I bounded or unbounded) defined by

$$(5) \quad u \longmapsto \mathcal{I}_h(u) := \int_I h(u(x)) dx,$$

where the integral is interpreted in the sense of Rockafellar (cf. [18, p. 7]). Thus \mathcal{I}_h is a well-defined operator from $L^1(I)$ to $[-\infty, \infty]$ and, since the entropy h is convex, also convex on $L^1(I)$.

For the conjugate of this integral functional, we take I to be a bounded interval and let $L := L^1(I)$ and $L^* := L^\infty(I)$. One can define a bilinear product on $L \times L^*$ by

$$(6) \quad (u, u^*) \longmapsto \langle u, u^* \rangle := \int_I u(x)u^*(x) dx.$$

To compute the convex conjugate of \mathcal{I}_h with respect to (6), we may conjugate the integrand, as in the following proposition.

PROPOSITION 1. *Let I be bounded and consider the pair $\langle L, L^* \rangle$ of subspaces of $L^1(I)$ as defined above with bilinear product (6). Then for any $q \in L^*$, we have*

$$\mathcal{I}_h^*(q) = \int_I h^*(q(t)) dt.$$

The proof of Proposition 1 can be found in either [19] or [16]. Finally, we recall a Fenchel duality theorem in its *partially finite* version. The proof of Theorem 1 as stated can be inferred from Theorem 4.2 in [4], with the attainment of the infimum proved via Theorems 3.7 and 3.8 of [7]. In what follows, “ri” denotes the relative interior of a subset of \mathbb{R}^n and “dom” denotes the effective domain of a convex function (i.e., the set of points at which the function is finite).

THEOREM 1. *Let V and V^* be vector spaces, and let $\langle \cdot, \cdot \rangle$ be a bilinear product on $V \times V^*$. Let $G: V \rightarrow \mathbb{R}^n$ be a linear map with adjoint G^T , let $F: V \rightarrow \bar{\mathbb{R}}$ be a proper convex function, and let $g: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ be a proper concave function. Then, under the CQ*

$$(CQ) \quad \text{ri}(G \text{ dom } F) \cap \text{ri}(\text{dom } g) \neq \emptyset,$$

we have

$$(7) \quad \inf \{ F(u) - g(Gu) \mid u \in V \} = \sup \{ g^*(\boldsymbol{\lambda}) - F^*(G^T \boldsymbol{\lambda}) \mid \boldsymbol{\lambda} \in \mathbb{R}^n \},$$

with the supremum on the right being attained when finite. Moreover for the case where V is a normed vector space with dual V^* , if F is strongly rotund (i.e., satisfies the three conclusions of Lemma 2), the infimum on the left is attained at a unique u .

3. Duality results. Theorem 1 directly applies to (\mathcal{P}) with I bounded. That is, $n = m + 1$, $V = L^1(I)$, $V^* = L^\infty(I)$, $\langle \cdot, \cdot \rangle$ is given by (6), $G := \mathbb{A}$, and the function g is defined by

$$g(\mathbb{A}p) = -\delta(\mathbb{A}p - \mathbf{d} \mid \mathbf{0}).$$

Lastly, $F = \mathcal{I}_h$, where h is the Boltzmann–Shannon entropy functional defined by (2), which by Lemma 2 is strongly rotund. The CQ amounts to

$$(8) \quad (CQ) \quad \mathbf{d} \in \text{ri}(\mathbb{A} \text{ dom } \mathcal{I}_h).$$

Precisely, we have the following.

COROLLARY 1. *Let I be bounded and assume (8) holds. Then (\mathcal{P}) has a unique solution and*

$$(9) \quad \begin{aligned} & \inf \{ \mathcal{I}_h(p) \mid p \in L^1(I), \mathbb{A}p = \mathbf{d} \} \\ & = \sup \left\{ \sum_{i=0}^m \lambda_i d_i - \mathcal{I}_h^*(\mathbb{A}^T(\boldsymbol{\lambda})) \mid \boldsymbol{\lambda} \in \mathbb{R}^{m+1} \right\}. \end{aligned}$$

Moreover the solution of the primal problem (left-hand side of (9)) is

$$(10) \quad e^{\sum_{i=0}^m \bar{\lambda}_i c_i(x)},$$

with $\bar{\lambda}_i$ being the unique solution to the dual problem (right-hand side of (9)).

It is straightforward to check (see section 5) that

$$\mathcal{I}_h^*(\mathbb{A}^T(\boldsymbol{\lambda})) = \int_I e^{\mu(x)} dx, \quad \mu(x) := \sum_{i=0}^m \lambda_i c_i(x),$$

where one can explicitly carry out the integration (cf. (16)). We also note in section 5 that the distribution given by (10) is indeed a probability distribution.

As previously mentioned, the case where $I = [0, \infty)$ is best treated differently. The duality result is identical; however, to prove it we shall bypass the direct application of Theorem 1 and exploit properties of the value function. We rewrite the CQ as

$$(11) \quad (\text{CQ}) \quad \mathbf{d} \in \text{ri } \mathcal{A},$$

where

$$\mathcal{A} := \{ \mathbf{x} \in \mathbb{R}^{m+1} \mid \exists p \in L^1[0, \infty) \text{ with } \mathcal{I}_h(p) \text{ finite and } \mathbb{A}p = \mathbf{x} \}.$$

We have the following theorem.

THEOREM 2. *Let $I = [0, \infty)$ and assume (11) holds. Then (\mathcal{P}) has a unique solution and*

$$(12) \quad \begin{aligned} & \inf \{ \mathcal{I}_h(p) \mid p \in L^1([0, \infty)), \mathbb{A}p = \mathbf{d} \} \\ &= \sup \left\{ \sum_{i=0}^m \lambda_i d_i - \int_0^\infty e^{\mu(x)} dx \mid \boldsymbol{\lambda} \in \mathbb{R}^{m+1} \right\}. \end{aligned}$$

Moreover, the solution of the primal problem is

$$(13) \quad e^{\sum_{i=0}^m \bar{\lambda}_i c_i(x)},$$

with $\bar{\lambda}_i$ being the unique value of the right-hand side of (12).

Proof of Theorem 2. Consider the value function

$$\mathcal{V}(\mathbf{d}) := \inf \{ \mathcal{I}_h(p) \mid \mathbb{A}p = \mathbf{d} \} = \inf \{ \mathcal{I}_h(p) + \delta(\mathbb{A}p - \mathbf{d} \mid \mathbf{0}) \mid p \in L^1([0, \infty)) \}.$$

We prove that under the (CQ) of (11),

$$(14) \quad \mathcal{V}(\mathbf{d}) = \sup \left\{ \sum_{i=0}^m \lambda_i d_i - \int_0^\infty e^{\mu(x)} dx \mid \boldsymbol{\lambda} \in \mathbb{R}^{m+1} \right\}.$$

First note that (14) easily holds with $=$ replaced with \geq . To see this, note that by the definition of h^* , for every $p \in \text{dom } \mathcal{I}_h$ with $\int c_i(x)p(x)dx = d_i$, we have

$$\int h^*(\mu)dx + \int h(p)dx \geq \int \sum_{i=0}^m \lambda_i c_i(x)p(x) dx = \sum_{i=0}^m \lambda_i d_i$$

holding for any $\boldsymbol{\lambda} \in \mathbb{R}^{m+1}$. The inequality follows by first taking the infimum over all such p , and then the supremum over $\boldsymbol{\lambda} \in \mathbb{R}^{m+1}$.

We now prove the reverse inequality. The (CQ) implies that $\mathbf{d} \in \text{ri}(\text{dom } \mathcal{V})$. Moreover, it is easily verified that \mathcal{V} is convex on its domain. Hence (see, for example, [3]), there exists a $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^{m+1}$ such that $\bar{\boldsymbol{\lambda}} \in \partial \mathcal{V}(\mathbf{d})$, the subgradient of \mathcal{V} at \mathbf{d} . Thus for all $\mathbf{z} \in \mathbb{R}^{m+1}$, $\mathcal{V}(\mathbf{z}) \geq \mathcal{V}(\mathbf{d}) + \langle \bar{\boldsymbol{\lambda}}, \mathbf{z} - \mathbf{d} \rangle$. Fix $M > 0$. Restricting our attention to p with support in $[0, M]$, we have (by definition of $\mathcal{V}(\mathbf{z})$) for all $p \in L^1([0, M])$

$$\mathcal{V}(\mathbf{d}) - \langle \bar{\boldsymbol{\lambda}}, \mathbf{d} \rangle \leq \mathcal{I}_h(p) - \langle \bar{\boldsymbol{\lambda}}, \mathbb{A}p \rangle.$$

Setting $\bar{\mu}(x) = \sum_{i=0}^m \bar{\lambda}_i c_i(x)$, we have

$$\mathcal{V}(\mathbf{d}) - \sum_{i=0}^m \bar{\lambda}_i d_i \leq \int_0^M (h(p(x)) - p(x) \bar{\mu}(x)) dx,$$

and hence

$$\sup_{p \in L^1[0, M]} \left\{ \int_0^M p(x) \bar{\mu}(x) - h(p(x)) dx \right\} \leq \sum_{i=0}^m \bar{\lambda}_i d_i - \mathcal{V}(\mathbf{d}).$$

The left-hand side of the above is by definition $\mathcal{I}_h^*(\bar{\mu})$. Hence applying Proposition 1 to $\mathcal{I}_h(p)$ on $[0, M]$, we have

$$\mathcal{I}_h^*(\bar{\mu}) = \int_0^M h^*(\bar{\mu}) dx = \int_0^M e^{\bar{\mu}(x)} dx \leq \sum_{i=0}^m \bar{\lambda}_i d_i - \mathcal{V}(\mathbf{d}),$$

or

$$\sum_{i=0}^m \bar{\lambda}_i d_i - \int_0^M e^{\bar{\mu}(x)} dx \geq \mathcal{V}(\mathbf{d}).$$

Since the above holds for each $M > 0$, the monotone convergence theorem implies

$$(15) \quad \sum_{i=0}^m \bar{\lambda}_i d_i - \int_0^\infty e^{\bar{\mu}(x)} dx \geq \mathcal{V}(\mathbf{d}).$$

Lastly, we prove primal attainment. The (CQ) holds, and hence the supremum on the right of (12) is finite, and moreover the previous analysis shows that there exists $\bar{\lambda}$ which attains this supremum. It remains to show that the dual function

$$D(\lambda) := \sum_{i=0}^m \lambda_i d_i - \int_0^\infty e^{\mu(x)} dx$$

is differentiable at $\lambda = \bar{\lambda}$. To this end, we note that by (15),

$$\int_{k_m}^\infty e^{\bar{\mu}(x)} dx < \infty.$$

Since for $x > k_m$, $\bar{\mu}(x) = \bar{\lambda}_0 + x \sum_{i=1}^m \bar{\lambda}_i - \sum_{i=1}^m k_i \bar{\lambda}_i$, we must have $\sum_{i=1}^m \bar{\lambda}_i < 0$, and hence $D(\lambda)$ is differentiable at $\lambda = \bar{\lambda}$. Thus

$$d_k = \int_0^\infty c_k(x) e^{\bar{\mu}(x)},$$

$\bar{p}(x) := e^{\bar{\mu}(x)}$ is feasible for the primal problem, and

$$\mathcal{I}_h(e^{\bar{\mu}(x)}) = \sum_{k=0}^m \bar{\lambda}_k d_k - \int_0^\infty e^{\bar{\mu}(x)} dx.$$

Since equality holds in (12), $e^{\bar{\mu}(x)}$ must indeed be the MES. The uniqueness follows from the strict convexity of the entropy (see, for example, [3]). \square

In the following sections we complement Corollary 1 and Theorem 2 by giving an explicit characterization of the (CQ) for our problem (\mathcal{P}), and by computing the dual function D explicitly in a form with no integrals.

4. The CQ. In Proposition 2 below, we give an explicit form of the CQ for problem (P), first for $I = [0, \infty)$ and then for $I = [0, K]$. We shall need the following simply lemma, whose proof is left as an exercise.

LEMMA 1. Let $I = [0, \infty)$ and $\varphi(x) := [\mathbb{A}^T \boldsymbol{\lambda}](x) = \lambda_0 + \lambda_1 c_1(x) + \dots + \lambda_m c_m(x)$. The following conditions are equivalent:

- (a) for all p s.t. $\mathcal{I}_h(p)$ is finite and $\mathbb{A}p \in \mathbb{R}^{m+1}$, we have $\langle \boldsymbol{\lambda}, \mathbb{A}p \rangle \geq 0$;
- (b) $\varphi(x) \geq 0$ for all $x \in \mathbb{R}_+$;
- (c) $M\boldsymbol{\lambda} \geq \mathbf{0}$ (componentwise), where

$$M := \begin{pmatrix} 1 & & & & & \\ 1 & k_2 - k_1 & & & & \\ \vdots & \vdots & \ddots & & & \\ 1 & k_m - k_1 & \cdots & k_m - k_{m-1} & & \\ 0 & 1 & \cdots & 1 & 1 & 1 \end{pmatrix}.$$

If A is an $(m \times n)$ -matrix and K_A is the convex cone defined by $K_A := \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} \geq \mathbf{0}\}$, one may easily verify that for the dual cone K_A^+ , we have

$$K_A^+ := \{\mathbf{y} \in \mathbb{R}^n \mid \langle \mathbf{y}, \mathbf{x} \rangle \geq 0 \forall \mathbf{x} \in K_A\} = A^T \mathbb{R}_+^m,$$

where A^T denotes the adjoint of A (for example, see [3]).

PROPOSITION 2. Let $I = [0, \infty)$ and $m > 2$. Then $(1, d_1, \dots, d_m)$ satisfies the CQ (11) for (P) if and only if $(d_1, \dots, d_m)^T$ satisfies

$$d_m > 0, \quad N^{-1}B(d_1, \dots, d_m)^T > \mathbf{0}, \quad \text{and} \quad \langle N^{-1}B(d_1, \dots, d_m)^T, \mathbf{u} \rangle < 1,$$

in which \mathbf{u} is the vector of appropriate dimension whose components are all equal to 1, and N and B are, respectively, the $(m - 1) \times (m - 1)$ - and $(m - 1) \times m$ -matrices given by

$$N := \begin{pmatrix} k_2 - k_1 & \cdots & k_m - k_1 \\ & \ddots & \vdots \\ & & k_m - k_{m-1} \end{pmatrix}, \quad B := \begin{pmatrix} 1 & & -1 \\ & \ddots & \vdots \\ & & 1 & -1 \end{pmatrix}.$$

Proof. We denote by cl the closure of a subset of \mathbb{R}^n . A classical separation argument shows that the vector $\mathbf{d}' \in \mathbb{R}^{1+m}$ does not belong to the closed convex set $\text{cl } \mathcal{A}$ if and only if there exists $\boldsymbol{\lambda} \in \mathbb{R}^{1+m}$ such that

- (α) $\langle \boldsymbol{\lambda}, \mathbf{d}' \rangle < 0$, and
- (β) $\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle \geq 0$ for all $\boldsymbol{\xi} \in \text{cl } \mathcal{A}$.

Clearly, $\text{cl } \mathcal{A}$ can be replaced by \mathcal{A} in condition (β), which can thus be rewritten as

$$(\beta') \langle \mathbb{A}^T \boldsymbol{\lambda}, p \rangle \geq 0 \text{ for all } p \text{ s.t. } \mathcal{I}_h(p) \text{ is finite and } \mathbb{A}p \in \mathbb{R}^{m+1}.$$

But from Lemma 1, the latter condition is equivalent to $M\boldsymbol{\lambda} \geq \mathbf{0}$. In other words, we have shown that $\mathbf{d}' \in \text{cl } \mathcal{A}$ if and only if for all $\boldsymbol{\lambda} \in \mathbb{R}^{1+m}$, either $\langle \boldsymbol{\lambda}, \mathbf{d}' \rangle \geq 0$ or $M\boldsymbol{\lambda} \not\geq \mathbf{0}$.

Let us define $C_M = \{\boldsymbol{\lambda} \in \mathbb{R}^{1+m} \mid M\boldsymbol{\lambda} \geq \mathbf{0}\}$. We have

$$\begin{aligned} \text{cl } \mathcal{A} &= \{\mathbf{d}' \mid \forall \boldsymbol{\lambda}, M\boldsymbol{\lambda} \not\geq \mathbf{0} \text{ or } \langle \boldsymbol{\lambda}, \mathbf{d}' \rangle \geq 0\} \\ &= \{\mathbf{d}' \mid \forall \boldsymbol{\lambda}, \boldsymbol{\lambda} \notin C_M \text{ or } \langle \boldsymbol{\lambda}, \mathbf{d}' \rangle \geq 0\} \\ &= \{\mathbf{d}' \mid \forall \boldsymbol{\lambda} \in C_M, \langle \boldsymbol{\lambda}, \mathbf{d}' \rangle \geq 0\} \\ &= C_M^+. \end{aligned}$$

By the previously mentioned characterization of C_M^+ as well as by standard properties of the relative interior of convex sets (see [17], section 6), we obtain

$$\text{ri } \mathcal{A} = \text{ri cl } \mathcal{A} = \text{ri } M^T \mathbb{R}_+^{1+m} = M^T \text{ri } \mathbb{R}_+^{1+m} = M^T(0, \infty)^{1+m}.$$

Consequently, $(1, d_1, \dots, d_m)$ belongs to $\text{ri } \mathcal{A}$ if and only if

$$\left\{ \begin{array}{l} 1 = \xi_0 + \xi_1 + \dots + \xi_{m-1} \\ d_1 = (k_2 - k_1)\xi_1 + \dots + (k_m - k_1)\xi_{m-1} + \xi_m \\ \vdots \\ d_{m-1} = (k_m - k_{m-1})\xi_{m-1} + \xi_m \\ d_m = \xi_m \end{array} \right.$$

for some $\boldsymbol{\xi} > \mathbf{0}$. By subtracting the last line from lines $2, \dots, m - 1$ in the above system, we see that $(1, d_1, \dots, d_m) \in \text{ri } \mathbb{A} \text{ dom } \mathcal{I}_h$ if and only if

$$d_m > 0, \quad N^{-1}B(d_1, \dots, d_m)^T > \mathbf{0}, \quad \text{and} \quad \langle N^{-1}B(d_1, \dots, d_m)^T, \mathbf{u} \rangle < 1.$$

Notice that N is invertible since $k_m > \dots > k_1$ by assumption. \square

For the case of bounded $I = [0, K]$, one can show Proposition 2 holds with the one modification of replacing B by

$$B_K := \begin{pmatrix} 1 & & -\frac{K - k_1}{K - k_m} \\ & \ddots & \vdots \\ & & 1 & -\frac{K - k_{m-1}}{K - k_m} \end{pmatrix}.$$

The proof of this is similar to that of Proposition 2.

5. Maximizing the dual function. Recall from Corollary 1 that under the CQ (8), the optimal value of (\mathcal{P}) is equal to the optimal value of the dual problem

$$(\mathcal{D}) \quad \max \left\{ D(\lambda_0, \boldsymbol{\lambda}) := \lambda_0 + \sum_{i=1}^m \lambda_i d_i - \mathcal{I}_h^*(\mathbb{A}^T(\lambda_0, \boldsymbol{\lambda})) \mid (\lambda_0, \boldsymbol{\lambda}) \in \mathbb{R}^{1+m} \right\}.$$

The formal adjoint \mathbb{A}^T of \mathbb{A} is readily computed as

$$\mathbb{A}^T((\lambda_0, \boldsymbol{\lambda})) = \langle (\lambda_0, \boldsymbol{\lambda}), (1, \mathbf{c}(\cdot)) \rangle.$$

By Proposition 1, we have

$$\begin{aligned} \mathcal{I}_h^*(\mathbb{A}^T(\lambda_0, \boldsymbol{\lambda})) &= \int_I h^*(\lambda_0 + \langle \boldsymbol{\lambda}, \mathbf{c}(x) \rangle) dx \\ &= \exp \lambda_0 \times \int_0^K \exp \left[\sum_{i=1}^m \lambda_i (x - k_i)^+ \right] dx \\ &= \exp \lambda_0 \times \sum_{j=1}^m \int_{k_j}^{k_{j+1}} \exp \left[\left(\sum_{i=1}^j \lambda_i \right) x - \sum_{i=1}^j \lambda_i k_i \right] dt \\ (16) \quad &= \exp \lambda_0 \times \sum_{j=1}^m \left(\exp(-\nu_j) \frac{\exp \mu_j k_{j+1} - \exp \mu_j k_j}{\mu_j} \right), \end{aligned}$$

in which $k_{m+1} := K$, $\nu_j := \sum_{i=1}^j \lambda_i k_i$, and $\mu_j := \sum_{i=1}^j \lambda_i$. The expression $\mu_j^{-1}(\exp \mu_j k_{j+1} - \exp \mu_j k_j)$ is understood to be $k_{j+1} - k_j$ when $\mu_j = 0$.

For the case $I = [0, \infty)$, Theorem 2 directly gave rise to the same dual function (with the integration carried out over the entire half line). In this case we have $\int_0^\infty e^{\mu(x)}$ equal to (16) with $k_{m+1} := +\infty$ and the understanding that $\exp(-\infty)$ is equal to zero.

We remark that $e^{-\lambda_0}$ can be taken to be

$$Z(\boldsymbol{\lambda}) := \int_0^\infty \exp \left[\sum_{i=1}^m \lambda_i (x - k_i)^+ \right] dx,$$

and hence the dual function to be maximized can be written in terms of $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$ as

$$(17) \quad \log Z(\boldsymbol{\lambda}) - \sum_{i=1}^m \lambda_i d_i.$$

In particular, the MES is indeed a probability distribution and has the form

$$\frac{1}{Z(\boldsymbol{\lambda})} e^{\sum_{i=1}^m \lambda_i c_i(x)},$$

with λ_i maximizing (17).

6. The value function. The *value* (or *objective*) function associated with problem (\mathcal{P}) is defined for $\mathbf{d} = (1, d_1, \dots, d_m)$ by

$$\mathcal{V}(\mathbf{d}) := \inf \{ \mathcal{I}_h(p) \mid \mathbf{d} = \mathbb{A}p \}.$$

While it is known that the value function is continuous on the interior of the CQ set, it is not in general on its closure. It turns out that if \mathcal{V} is lower semicontinuous on its domain (the set of all feasible data), then there is no duality gap, i.e., (9) and (12) hold whenever the left-hand side is finite.

We will prove \mathcal{V} is lower semicontinuous on its domain for the case $I = [0, \infty)$. The proof for bounded I follows verbatim from the first part of the proof. Our proof of lower semicontinuity only requires the entropy functional (over a bounded domain) to have weakly compact level sets. The following result from [7] (Theorem 3.8) implies that our proof holds not just for h but also for any entropy whose convex conjugate is everywhere finite and differentiable.

LEMMA 2. *Let I be bounded and let $\phi : \mathbb{R} \rightarrow \bar{\mathbb{R}}$ be such that ϕ^* is everywhere finite and differentiable; then*

$$I_\phi(p) = \int_I \phi(p(x)) dx$$

(i) *is strictly convex, (ii) has weakly compact level sets in $L^1(I)$, and (iii) $p_n \rightarrow p$ in $L^1(I)$ whenever $I_\phi(p_n) \rightarrow I_\phi(p)$ and $p_n \rightarrow p$ weakly in $L^1(I)$.*

We will also need the following useful lemma, which explicitly gives the MES for the case of two constraints.

LEMMA 3. *The two-constraint problem,² i.e.,*

$$\left| \begin{array}{l} \text{minimize } \mathcal{I}_h(p) := \int_0^\infty h(p(x)) dx \\ \text{s.t. } d_0 = \int_0^\infty p(x) dx, \\ d_1 = \int_0^\infty x p(x) dx, \end{array} \right.$$

has the explicit solution

$$\hat{p}(x) = \frac{d_0^2}{d_1} e^{-(d_0/d_1)x}.$$

Proof of Lemma 3. Let $\lambda_0 = \log \frac{d_0^2}{d_1}$ and $\lambda_1 = -\frac{d_0}{d_1}$. One readily checks that $\hat{p}(x)$ is feasible (satisfies the two constraints), and $\mathcal{I}_h(\hat{p}) = d_0 \log d_0^2/d_1 - 2d_0$. On the other hand,

$$D(\lambda_0, \lambda_1) = \lambda_0 d_0 + \lambda_1 d_1 - \int_0^\infty e^{\lambda_0 + \lambda_1 x} dx = d_0 \log d_0^2/d_1 - 2d_0.$$

The result follows by (12)—in fact, the result would follow simply from weak duality, i.e., (12) with equality replaced by \geq , which always holds true. \square

THEOREM 3. *The value function \mathcal{V} is lower semicontinuous on its domain.*

Proof. The basis for our proof lies in the fact that the particular structure of the constraint functions allows us to rewrite all but the first two constraints as integrals over a finite domain. To this end, observe that for $j = 2, \dots, m$, we have

$$\begin{aligned} d_j &= \int_{k_j}^\infty (x - k_j)p(x) dx \\ (18) \quad &= \int_0^\infty xp(x) dx - k_j \int_0^\infty p(x) dx + \int_0^{k_j} (k_j - x)p(x) dx \\ &= d_1 - k_j + \int_0^{k_j} (k_j - x)p(x) dx. \end{aligned}$$

Consequently, all constraints corresponding to $j > 1$ can be rewritten as

$$\int_0^M (k_j - x)^+ p(x) dx = \delta_j := d_j - d_1 + k_j,$$

where M is any constant greater than or equal to k_m .

With this in hand, suppose $\mathbf{d}, \mathbf{d}^{(n)} \in \text{dom } \mathcal{V}$ ($d_0 = d_0^{(n)} = 1$) with $\mathbf{d}^{(n)} \rightarrow \mathbf{d}$ and for some constant C , $\mathcal{V}(\mathbf{d}^{(n)}) \leq C$ for all n . We prove that $\mathcal{V}(\mathbf{d}) \leq C$. To this end, pick a sequence $p^{(n)}$ such that $\mathbb{A}p^{(n)} = \mathbf{d}^{(n)}$ and $\mathcal{I}_h(p^{(n)}) \leq C + 2^{-n}$. Fix $M > k_m$ and define

$$d_{M,0}^{(n)} = \int_0^M p^{(n)}(x) dx \quad \text{and} \quad d_{M,1}^{(n)} = \int_0^M x p^{(n)}(x) dx.$$

²This constrained problem is used as a tool in our analysis. In the context of options, not only would $d_0 = 1$, but d_1 would also be predetermined by the risk-free interest rate.

Then $1 = d_{M,0}^{(n)} + \varepsilon_{M,0}^{(n)}$ and $d_1^{(n)} = d_{M,1}^{(n)} + \varepsilon_{M,1}^{(n)}$, where

$$\varepsilon_{M,0}^{(n)} = \int_M^\infty p^{(n)}(x) dx \quad \text{and} \quad \varepsilon_{M,1}^{(n)} = \int_M^\infty x p^{(n)}(x) dx.$$

Clearly, $0 \leq d_{M,0}^{(n)} \leq 1$ and $0 \leq d_{M,1}^{(n)} \leq d_1^{(n)} \rightarrow d_1$, so we have, up to taking a subsequence (not relabeled), that $d_{M,0}^{(n)}$ tends to some $d_{M,0}$ and $d_{M,1}^{(n)}$ tends to some $d_{M,1}$. Then $\varepsilon_{M,0}^{(n)} \rightarrow \varepsilon_{M,0} := 1 - d_{M,0}$ and $\varepsilon_{M,1}^{(n)} \rightarrow \varepsilon_{M,1} := d_1 - d_{M,1}$.

Assume for the moment that for some constant c ,

$$(19) \quad \int_0^M h(p^{(n)}(x)) dx < c.$$

Since h^* is everywhere finite and differentiable, Lemma 2 implies that there exists a subsequence (not relabeled) such that $p^{(n)}$ weakly converges to some p_M on $[0, M]$. Furthermore, p_M satisfies

$$\int_0^M (k_j - x)^+ p_M(x) dx = \delta_j, \quad j > 1,$$

$$\int_0^M p_M(x) dx = d_{M,0} \leq 1,$$

$$\int_0^M x p_M(x) dx = d_{M,1} \leq d_1.$$

We note that either

- (a) $d_{M,0} < 1$ and $d_{M,1} < d_1$, or
- (b) $d_{M,0} = 1$ and $d_{M,1} = d_1$.

For case (a), we consider the two-constraint problem

$$\left| \begin{array}{l} \text{minimize} \quad \mathcal{I}_h(\tilde{p}) := \int_M^\infty h(\tilde{p}(x)) dx \\ \text{s.t.} \quad \varepsilon_{M,0}^{(n)} = \int_M^\infty \tilde{p}(x) dx, \\ \quad \quad \varepsilon_{M,1}^{(n)} = \int_M^\infty x \tilde{p}(x) dx. \end{array} \right.$$

By Lemma 3, this has an *explicit* solution

$$\tilde{p}^{(n)}(x) = \frac{(\varepsilon_{M,0}^{(n)})^2}{\varepsilon_{M,1}^{(n)} - M\varepsilon_{M,0}^{(n)}} e^{-\frac{\varepsilon_{M,0}^{(n)}}{\varepsilon_{M,1}^{(n)} - M\varepsilon_{M,0}^{(n)}}(x-M)}.$$

Note that on $[M, \infty)$ the entropy of $\tilde{p}^{(n)}$ is

$$(20) \quad \varepsilon_{M,0}^{(n)} \log \left(\frac{(\varepsilon_{M,0}^{(n)})^2}{\varepsilon_{M,1}^{(n)} - M\varepsilon_{M,0}^{(n)}} \right) - 2\varepsilon_{M,0}^{(n)},$$

which, since $\varepsilon_{M,0}^{(n)}, \varepsilon_{M,1}^{(n)}$ are bounded, is bounded below. Moreover, $\tilde{p}^{(n)}(x)$ converges pointwise to

$$\tilde{p}(x) := \frac{\varepsilon_{M,0}^2}{\varepsilon_{M,1} - M\varepsilon_{M,0}} e^{-\frac{\varepsilon_{M,0}}{\varepsilon_{M,1} - M\varepsilon_{M,0}}(x-M)}.$$

Note that for case (a), $\varepsilon_{M,1} - M\varepsilon_{M,0} > 0$. Define \hat{p} to be p_M on $[0, M]$ and \tilde{p} on $[M, \infty)$. Then \hat{p} is feasible for \mathbf{d} , and by taking a subsequence (not relabeled) of $p^{(n)}$, we have

$$\begin{aligned} \mathcal{I}_h(\hat{p}) &= \int_0^M h(p_M)dx + \int_M^\infty h(\tilde{p})dx \\ &\leq \int_0^M h(p^{(n)})dx + \int_M^\infty h(\tilde{p}^{(n)})dx + 2^{-n} \\ &\leq \int_0^\infty h(p^{(n)}) + 2^{-n} \leq C + 2^{-n} + 2^{-n}. \end{aligned}$$

Above we used the weak lower semicontinuity of \mathcal{I}_h on $[0, M]$ in the first inequality, and the fact that $\tilde{p}^{(n)}$ was optimal with respect to its constraints on $[M, \infty)$ in the second inequality. Letting $n \rightarrow \infty$ gives $\mathcal{V}(\mathbf{d}) \leq C$.

In case (b), p_M (extended to be 0 on $[M, \infty)$) is feasible for \mathbf{d} . We have $p^{(n)} \rightarrow 0$ in L^1 on $[M, \infty)$, but since we do not know that I_h is lower semicontinuous on the infinite domain, we cannot immediately conclude anything about the limit of $\int_M^\infty h(p^{(n)})$. It suffices to prove that the liminf $\int_M^\infty h(p^{(n)}) = A$, with A for some finite $A \geq 0$. Then, by weak lower semicontinuity of \mathcal{I}_h on $[0, M]$, we may pick a subsequence to find

$$\mathcal{V}(\mathbf{d}) \leq \mathcal{I}_h(p_M) \leq \int_0^M h(p_M) dx + A \leq \int_0^\infty h(p^{(n)})dx + 2^{-n} \leq C + 2^{-n+1}.$$

To this end, we note that since $\mathcal{I}_h(p^{(n)}) < C + 2^{-n}$ and $\int_0^M h(p^{(n)})$ is bounded below, $\liminf \int_M^\infty h(p^{(n)})$ cannot be $+\infty$. Moreover, since both $\varepsilon_{M,0}^{(n)}$ and $\varepsilon_{M,1}^{(n)}$ tend to 0, the liminf of the entropies of the optimal $\tilde{p}^{(n)}$ (i.e., (20)) is greater than or equal to zero. Since $p^{(n)}$ restricted to $[M, \infty)$ always has greater entropy than $\tilde{p}^{(n)}$, $\liminf \int_M^\infty h(p^{(n)})$ is some finite number $A \geq 0$.

Finally we address assumption (19). Suppose this did not hold; then (up to taking a subsequence) the entropy of $p^{(n)}$ on $[M, \infty)$ would have to approach $-\infty$. But this is impossible since we have shown above that the optimal (lowest entropy) distribution on $[M, \infty)$, over constraints for which $p^{(n)}$ restricted to $[M, \infty)$ is admissible, has entropy bounded below. \square

COROLLARY 2. *Equality holds in (9) and (12) whenever the left-hand side is finite.*

Proof. See [18]. \square

7. Remark. We briefly comment on the presence of a prior distribution. For a fixed distribution q (i.e., $q \in L^1(I)$, $\int_I q(x) dx = 1$), consider

$$(\mathcal{P}_q) \quad \left| \begin{array}{l} \text{minimize} \quad \int_I p(x) \log \left(\frac{p(x)}{q(x)} \right) dx \\ \text{s.t.} \quad 1 = \int_I p(x) dx \quad \text{and} \quad d_j = \int_I c_j(x)p(x) dx. \end{array} \right.$$

Here we minimize the “entropic” distance to a prior distribution $q(x)$. This gives a more realistic approach to recovering the price distribution, as our previous model

is based upon the assumption that the only a priori guess for $p(x)$ is uniform. In practice, one may have a priori information that the unknown distribution could be, say, log-normal.

For the analysis to carry over, we require q to be bounded away from zero at $x = 0$. Particularly, we would require

$$e^{-ax} < q(x) < e^{bx} \quad \text{a.e. for some positive constants } a, b.$$

This assumption may seem rather odd but it is simply a consequence of the structure of MESs. Note, for example, that the MES is never zero when $x = 0$ regardless of the moment constraints.

The main modification in the results would be that the measure dx in the dual function D is replaced with $q(x)dx$, with the corresponding adjustment in the closed form of the primal solution. Note that this would prevent one from carrying out the integration performed in (16) for an explicit representation. In this way, the uniform prior is rather special.

REFERENCES

- [1] M. AVELLANEDA, *The minimum-entropy algorithm and related methods for calibrating asset-pricing models*, in Proceedings of the International Congress of Mathematicians, Vol. III, Doc. Math., Berlin, 1998, pp. 545–563.
- [2] M. AVELLANEDA, C. FRIEDMAN, R. HOLMES, AND D. SAMPERI, *Calibrating volatility surfaces via relative entropy minimization*, Appl. Math. Finance, 4 (1997), pp. 37–64.
- [3] J. M. BORWEIN AND A. S. LEWIS, *Convex Analysis and Nonlinear Optimization: Theory and Examples*, CMS Books Math./Ouvrages Math. SMC 3, Springer-Verlag, New York, 2000.
- [4] J. M. BORWEIN AND A. S. LEWIS, *Partially finite convex programming. Part I: Quasi relative interiors and duality theory*, Math. Programming, 57 (1992), pp. 15–48.
- [5] J. M. BORWEIN AND A. S. LEWIS, *Partially finite convex programming. Part II: Explicit lattice models*, Math. Programming, 57 (1992), pp. 49–83.
- [6] J. M. BORWEIN AND A. S. LEWIS, *Partially-finite programming in L_1 and the existence of maximum entropy estimates*, SIAM J. Optim., 3 (1993), pp. 248–267.
- [7] J. M. BORWEIN AND A. S. LEWIS, *Strong rotundity and optimization*, SIAM J. Optim., 4 (1994), pp. 146–158.
- [8] J. M. BORWEIN AND M. A. LIMBER, *Underdetermined Moment Problems: A Case for Convex Analysis*, invited, SIAM Conference on Optimization, SIAM, Philadelphia, 1994.
- [9] P. W. BUCHEN AND M. KELLEY, *The maximum entropy distribution of an asset inferred from option prices*, J. Financial and Quantitative Analysis, 31 (1996), pp. 143–159.
- [10] J. COVER AND J. A. THOMAS, *Elements of Information Theory*, John Wiley, New York, 1991.
- [11] J. B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms. Vol. I: Fundamentals*, Springer-Verlag, Berlin, 1993.
- [12] J. B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms. Vol. II: Advanced Theory and Bundle Methods*, Springer-Verlag, Berlin, 1993.
- [13] J. C. HULL, *Options, Futures, and Other Derivative Securities*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [14] D. M. LIN AND E. K. WONG, *A survey on the maximum entropy method and parameter spectral estimation*, Phys. Rep., 193 (1990), pp. 41–135.
- [15] P. MARÉCHAL, *On the principle of maximum entropy as a methodology for solving linear inverse problems*, in Probability Theory and Mathematical Statistics, B. Grigelionis et al., eds., VPS/TEV, Zeist, The Netherlands, 1999, pp. 481–492.
- [16] P. MARÉCHAL, *A note on entropy optimization*, in Approximation, Optimization and Mathematical Economics, M. Lassonde, ed., Physica-Verlag, Heidelberg, 2001, pp. 205–211.
- [17] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [18] R. T. ROCKAFELLAR, *Conjugate Duality and Optimization*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 16, SIAM, Philadelphia, 1974.
- [19] R. T. ROCKAFELLAR, *Convex integral functionals and duality*, in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, New York, 1971, pp. 215–236.