15th Triennial World Congress, Barcelona, Spain

BOUNDED UNCERTAINTY MODELS IN FINANCE: PARAMETER ESTIMATION AND FORECASTING

Laurent El Ghaoui * **and Giuseppe Calafiore** **,[1]

* *Department of Electrical Engineering and Computer Science, UC Berkeley, CA. e-mail:* `elghaoui@eecs.berkeley.edu`
** *Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy. e-mail:* `calafiore@polito.it`

Abstract: In this paper we consider the problem of modelling observed data using a class of multivariate models with unknown-but-bounded (ubb) noise and uncertainty. Standard ARX models with additive and multiplicative bounded noise belong to the considered class, as well as the deterministic counterpart of ARCH models extensively used in econometrics. We outline a method to fit these models based on historical data, and discuss the issues of set-valued forecasting.

Keywords: Identification, Bounded uncertainty, Semidefinite programming, Financial modelling.

1. INTRODUCTION

1.1 Uncertainty models

Models with bounded uncertainty, or *uncertain systems* for short, have been around for a long time in the systems and control community, (Boyd *et al.*, 1994; Doyle, 1982). Reasons for their success in control applications can be summarized as follows. These models lend naturally themselves to worst-case (or very rare event) analysis. Bounded uncertainty models do not assume any prior information on the distribution of the uncertain parameters. Very efficient approximation methods, mostly based on convex optimization, can be devised for computing prediction bounds; by "efficient" we mean that these algorithms scale extremely well when the model size (for instance, the number of assets under consideration) grows. Similarly, efficient convex optimization methods for fitting these models based on historical data have been (recently) introduced. Finally, these models can be mixed with stochastic description. For example, it is possible to consider stochastic models with deterministic uncertainty, such as Brownian motions with unknown-but-bounded volatility.

In this paper, we explore some of the ideas involved in uncertainty models, and discuss their relevance to finance applications. In particular, we introduce a new class of multivariate bounded uncertainty models that capture some of the features of the celebrated AR/ARCH models.

1.2 Basic idea

The basic idea behind deterministic uncertainty models is as follows. Consider a simple stochastic model for an asset price $S(k)$

$$S(k+1) = (\gamma_0 + \sigma\epsilon(k))S(k),$$

[1] G. Calafiore acknowledges funding from the CNR Agenzia 2000 program.

where γ_0, σ are constants, and $\epsilon(k)$ is a standard Gaussian sequence, see for instance (Luenberger, 1995). A bounded-uncertainty counterpart of the above model would be

$$S(k+1) = (\gamma_0 + \sigma u(k))S(k), \quad \|u(k)\|_2 \leq 1.$$

The above system may be viewed as a deterministic system subject to a time-varying input u that is unknown-but-bounded. The prediction problem for the stochastic process is to compute the distribution of the asset price at some future time T. For the deterministic counterpart, the prediction problem is to compute bounds on the asset price.

It can be argued that, since we are computing the bounds on the asset price, the resulting prediction will be overly conservative. This can be true if one is not careful about the choice of the model. However there are several arguments in favor of bounded-uncertainty models. First, in some situations, the worst-case is really what we are trying to predict. This notion is useless for stochastic models with non-compactly supported distributions, but for uncertainty models it does make sense. Second, the uncertainty models should be useful in the context of *approximation*. The prediction problems for general uncertainty models is very difficult, but there are systematic approximation methods that are very efficient. In short, these models seem to be much more computationally tractable than stochastic models.

As said before, one should not make a conflict out of the distinction between stochastic and deterministic uncertainty, since it is possible to consider models with both kinds of uncertainty. Such models have been considered for instance by (Avellaneda and Parás, 1996), but their generalization to multivariate situations is out of the scope of this paper.

In this paper, we will exhibit multivariate uncertainty models that are deterministic counterparts to AR/ARCH models. We also discuss how the prediction problem can be attacked. General issues related to non-probabilistic estimation may be found in the classical references (Norton and Veres, 1991), (Kurzhanski and Valyi, 1996).

Notation

For a square matrix X, $X \succ 0$ (resp. $X \succeq 0$) means X is symmetric, and positive-definite (resp. semidefinite). For $P \in \mathbf{R}^{n\times n}$, with $P \succ 0$, and $x \in \mathbf{R}^n$, the notation $\mathcal{E}(P, x)$ denotes the ellipsoid

$$\mathcal{E}(P, x) = \left\{\xi : (\xi - x)^T P^{-1}(\xi - x) \leq 1\right\},$$

where x is the center, and P determines the "shape" of the ellipsoid.

2. AR/ARCH MODELS

2.1 Stochastic GARCH models

GARCH (Generalized AutoRegressive Conditional Heteroskedastic) models are stochastic models introduced by (Bollerslev, 1986) for modelling financial data. If $S(k)$ denotes a price (of a share, a bond, etc), the AR(r)/GARCH(p,q) model describes the evolution of the log-return

$$x(k) = \log \frac{S(k)}{S(k-1)},$$

in terms of the stochastic difference equation

$$x(k) = \gamma_0 + \sum_{l=1}^{r} \gamma_l x(k-l) + \sigma(k)\epsilon(k),$$

where $\epsilon(k)$ is a standard Gaussian variable, and the "volatility" $\sigma(k)$ follows the dynamics

$$\sigma(k)^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i x(k-i)^2 + \sum_{j=1}^{q} \beta_j \sigma(k-i)^2.$$

For a discussion of these models, and their relevance in a financial context, see for instance (Shiryaev, 1999). Multivariate versions of GARCH models are discussed in (Gouriéroux, 1997).

To fit these models, that is, compute the parameters α_i, $1 \leq i \leq p$, β_j, $1 \leq j \leq q$, and γ_l, $1 \leq l \leq r$, given historical values of the returns $x(k)$, $1 \leq k \leq N$, we can use the maximum likelihood method. In the case of an AR(r)/GARCH(p,0) model, we must maximize the function

$$p_{\alpha,\gamma}(x) = \Pi_{k=1}^{n}(\alpha_0 + \sum_{i=1}^{p} \alpha_i x(k)^2)^{-1/2} \times$$
$$\exp\left\{-\frac{1}{2}\sum_{k=1}^{N} \frac{(x(k) - \gamma_0 - \sum_{l=1}^{r} \gamma_l x(k-l))^2}{\alpha_0 + \sum_{i=1}^{p} \alpha_i x(k)^2}\right\}.$$

The problem of maximizing the log of this function is a convex programming problem of the MAXDET type, that can be solved in polynomial time (Vandenberghe *et al.*, 1998).

2.2 Uncertain Vector AR/ARCH Models

Since the appearance of GARCH models, a large number of variants have been proposed, including multivariate versions. We will not discuss these variants here. Instead, we introduce a multivariate, *bounded uncertainty* counterpart to the AR/ARCH models.

We define the model as follows:

$$x(k) = Ax(k-1) + b + P(k)^{1/2}u(k), \quad (1)$$

where $\|u(k)\|_2 \leq 1$, and $A \in \mathbf{R}^{n\times n}$, $b \in \mathbf{R}^n$ are given, and the vector $u(k)$ is unknown-but-bounded. The positive semidefinite matrix $P(k)$, the square root of which plays the role of a volatility matrix, obeys the following recurrence relation

$$P(k) = P + \sum_{i=1}^{p} \alpha_i Q(k-i), \qquad (2)$$

where $P \succeq 0$ is a positive semidefinite matrix, $Q(k-i) \doteq x(k-i)x(k-i)^T$, and the α_i's are positive numbers. We note that the recurrence relation is well-defined, in the sense that the right-hand side in the above equation is a positive semidefinite matrix.

Let us check that the above model captures some of the structure of the classical AR/ARCH model. Consider the univariate case ($n = 1$ and x is scalar). We observe that the new iterate is an affine combination of the previous return $x(k-1)$ and a noise term. The noise term $P(k)^{1/2}u(k)$ has two components: $u(k)$ has the same role as the Gaussian noise $\epsilon(k)$ in the stochastic model. The other component $P(k)^{1/2}$ is a deterministic equivalent to the variance. Its square $P(k)$ obeys a recurrence relation that is affine in the previous values of the squares, $x(k-i)^2$, $1 \leq i \leq p$.

3. PARAMETER FITTING

We now examine the problem of fitting the parameters of the above model, namely A, b, P, α, to data. In the parallel work (Calafiore *et al.*, 2002), we are developing a general statistical theory of identification of set-valued models based on consistency criteria, analogous to the one discussed in this paper.

Assume we are given historical returns $x(k)$ over a time period: $1 \leq k \leq N$. We define the vector θ as the collection of $m := 3n(n+1)/2 + p + q$ parameters needed to define the model, in short $\theta = (A, b, P, \alpha)$.

Our approach is as follows: we require that, at each step k, $1 \leq k \leq N$, the recurrence equations (1-2) should be consistent, in the sense that for every k, $1 \leq k \leq N$, there exists a vector $u(k)$ with $\|u(k)\|_2 \leq 1$, such that the equations (1-2) hold. These consistency relations will impose a number of constraints on the model variable θ. We will observe that the constraints define a *convex* set $\mathcal{C}$, that has the form

$$\mathcal{C} = \left\{ \theta \mid F(\theta) = F_0 + \sum_{i=1}^{m} \theta_i F_i \succeq 0 \right\}, \qquad (3)$$

where $F_0, \ldots, F_m$ are symmetric matrices that depend on the historical data only, and the notation $F(\theta) \succeq 0$ means that $F(\theta)$ is positive semidefinite. It is easily verified that the above set is convex.

Once the variables are constrained in the set $\mathcal{C}$ by the consistency relations, it suffices to select a member in the set to serve as our "best estimate" for the parameters. There are of course an infinite number of possible choices. We will take a specific solution in this set, by minimizing a linear objective under the constraint $F(\theta) \succeq 0$. The problem is a semidefinite program, solvable in polynomial-time.

We now turn to a more precise description of the set $\mathcal{C}$. For every k, $1 \leq k \leq N$, equation (1) holds for some $u(k)$, $\|u(k)\|_2 \leq 1$ if and only if

$$P(k) \succeq q(k)q^T(k),$$

where $q(k) \doteq Ax(k-1) + b - x(k)$. Equivalently, using the Schur complements rule the above is expressed as

$$\begin{bmatrix} P(k) & q(k) \\ q^T(k) & 1 \end{bmatrix} \succeq 0,$$

where

$$P(k) = P + \sum_{i=1}^{p} \alpha_i Q(k-i).$$

We see that the consistency equations are satisfied for some vectors $u(k)$, $\|u(k)\|_2 \leq 1$, if and only if for $1 \leq k \leq N$

$$F_k(\theta) := \begin{bmatrix} P + \sum_{i=1}^{p} \alpha_i Q(k-i) & q(k) \\ q^T(k) & 1 \end{bmatrix} \succeq 0.$$

The above conditions are convex in $\theta = (P, A, b, \alpha)$. In fact, we may describe the above constraints via a linear matrix inequality (LMI) description such as (3). Specifically, if we define the block-diagonal matrix

$$F(\theta) := \mathbf{diag}\left(F_1(\theta), \ldots, F_N(\theta), \alpha_1, \ldots, \alpha_p\right),$$

then the parameter $\theta = (A, b, P, \alpha)$ is consistent with observed data if and only if $F(\theta) \succeq 0$.

Our estimate is chosen by solving the following semidefinite program

$$\text{minimize } \mathbf{Tr} P + \sum_{i=1}^{p} \alpha_i \text{ subject to } F(\theta) \succeq 0.$$

The rationale behind the choice of the above objective function is to select, among all consistent models in the considered class, the one with smallest "covariance" $P(k)$.

4. PREDICTION

We now turn to the problem of predicting future asset prices, based on the model (1-2). We assume that $x(0)$ is given, and we seek to predict $x(T)$ for some future time instant $T \geq 0$.

In a stochastic setting, prediction means computing the distribution of the variable $x(T)$. In a deterministic uncertainty setting, the prediction problem is to compute the *reachable set*, that is, the set of states that are reachable by the system at time T. For general uncertainty models, this problem is in general extremely difficult. Hence the idea of computing *outer approximations* for reachable sets, that results in *guaranteed bounds* for $x(T)$, and related payoff functions.

To compute bounds in a recursive and tractable manner, one may use *ellipsoids of confidence* for the states. The approximate prediction problem is to recursively compute ellipsoids of confidence $\mathcal{E}_k := \mathcal{E}(X(k), \hat{x}(k))$ for the state at time k, for $0 \leq k \leq T$. This approach has been introduced in (El Ghaoui and Calafiore, 1999), for approximating the reachable set of uncertain linear systems. Here, $X(k)$ is a positive-definite matrix that determines the shape of the ellipsoid, and $\hat{x}(k)$ is its center.

4.1 Problem setup

Let us assume we computed ellipsoids of confidence $\mathcal{E}_{k-i} = \mathcal{E}(X(k-i), \hat{x}(k-i))$ for $1 \leq i \leq p$. Consider the problem of computing an ellipsoid of confidence for the new state,

$$x(k) = Ax(k-1) + b + \left(P + \sum_{i=1}^{p} \alpha_i Q(k-i)\right)^{1/2} u(k), \ \|u(k)\|_2 \leq 1. \tag{4}$$

We seek $X(k) \succ 0$ and $\hat{x}(k)$ such that

$$(x(k) - \hat{x}(k))^T X(k)^{-1} (x(k) - \hat{x}(k)) \leq 1 \tag{5}$$

whenever $x(k-i)$ belongs to the ellipsoid of confidence $\mathcal{E}_{k-i}$, $1 \leq i \leq p$

$$(x(k-i) - \hat{x}(k-i))(x(k-i) - \hat{x}(k-i))^T \preceq X(k-i), \tag{6}$$

and $x(k)$ satisfies (4) for some $u(k)$, $\|u(k)\|_2 \leq 1$, that is

$$P + \sum_{i=1}^{p} \alpha_i Q(k-i) \succeq q(k) q^T(k). \tag{7}$$

The above problem (that is, checking if the above is true) is hard in general.

4.2 Lagrange relaxation: basic idea

We are now going to apply a Lagrange relaxation method. The basic idea can be understood from the following simple example. Assume we seek to find a *sufficient* condition ensuring

$$\xi^T F \xi \leq 0$$

whenever

$$\phi(\xi\xi^T) \preceq 0,$$

where $\xi \in \mathbf{R}^m$ is a vector of variables, F is a symmetric $m \times m$ matrix, and ϕ is a linear mapping from the space of symmetric $m \times m$ matrices to the space of symmetric $N \times N$ matrices. The above problem is hard in general, but a simple sufficient condition is as follows. If there exists a symmetric, positive semidefinite $N \times N$ matrix $S \succeq 0$ such that

$$\text{for every } \xi \in \mathbf{R}^m, \ \ \xi^T F \xi \leq \mathbf{Tr} S \phi(\xi\xi^T),$$

then our original condition is true. For any given S, the above condition is a simple scalar condition on ξ

$$\xi^T F \xi \leq \xi^T \phi^*(S) \xi,$$

where ϕ^* is the map dual to ϕ. This scalar condition is equivalent to

$$F \preceq \phi^*(S),$$

which is a linear matrix inequality (LMI) condition on $S \succeq 0$. The morale of this example is as follows. A sufficient condition for a quadratic function to be negative whenever some quadratic matrix inequality holds, can be formulated in terms of a linear matrix inequality in some "Lagrange multiplier" matrix S. The latter is easily checked using semidefinite programming. We notice that the main tool used here (Lagrange relaxation for quadratic programming) is the same as the one used in the context of combinatorial optimization with astounding success, see for instance (Goemans and Williamson, 1994).

In our context, we will obtain a linear matrix inequality condition on some matrix $S \succeq 0$ that guarantees that the ellipsoid $\mathcal{E}(X(k+1), \hat{x}(k+1))$ is an ellipsoid of confidence for the new state. We will observe that this condition is *jointly* convex in S and the variables $X(k+1)$, $\hat{x}(k+1)$. We will use this fact to optimize the "size" of the new ellipsoid, via semidefinite programming.

4.3 LMI update conditions

We return to the problem of checking if inequality (5) holds whenever inequalities (7) and (6) hold. This condition is true if there exist Lagrange multiplier matrices $S(k) \succeq 0$ and $T(k-i) \succeq 0$ $(1 \le i \le p)$ such that

$$\delta^T(k)^T X(k)^{-1}\delta(k) \le 1 + \sum_{i=1}^{p} \mathbf{Tr} T(k-i) \left(\delta(k-i)\delta^T(k-i) - X(k-i)\right) + \mathbf{Tr} S(k) \left(q(k)q^T(k) - P - \sum_{i=1}^{p} \alpha_i Q(k-i) \right),$$

where we set $\delta(k) \doteq x(k) - \hat{x}(k)$. For fixed $S(k)$, $T(k-i)$, $\hat{x}(k)$ and $X(k+1)$, the above condition is a single scalar condition on the vector

$$\xi = \begin{bmatrix} 1 \\ x(k) \\ x(k-1) \\ \vdots \\ x(k-p) \end{bmatrix}.$$

We can express this condition as

$$\xi^T F(X(k), \hat{x}(k), S(k), T(k-1), \dots, T(k-p))\xi \le 0, \ \forall \xi,$$

for some appropriate symmetric matrix $F(\cdot)$ of size $np+1$. The above condition is equivalent to

$$F(X(k), \hat{x}(k), S(k), T(k-1), \dots, T(k-p)) \preceq 0.$$

We already know that the above is an LMI condition on $S(k), T(k-1), \dots, T(k-p)$ for fixed $X(k), \hat{x}(k)$. We will show that the above is in fact an LMI in all variables $X(k), \hat{x}(k), S(k), T(k-1), \dots, T(k-p)$. This is the consequence of simple but tedious algebra: The LMI condition on $X(k), \hat{x}(k), S(k), T(k-1), \dots, T(k-p)$ is

$$\begin{bmatrix} X(k) & H(\hat{x}(k)) & 0 \\ * & G_1(\cdot) & G_2(\cdot) \\ * & * & K(\cdot) \end{bmatrix} \succeq 0, \qquad (8)$$

where

$$H(\hat{x}(k)) \doteq \begin{bmatrix} \hat{x}(k) & I & 0 \end{bmatrix},$$

$$G_1(S(k), T(k-1), \dots, T(k-p)) \doteq \begin{bmatrix} \Omega(\cdot) & 0 & -\hat{x}(k-1)^T T(k-1) \\ * & 0 & 0 \\ * & * & T(k-1) - \alpha_1 S(k) \end{bmatrix} + \begin{bmatrix} b^T \\ -I \\ A^T \end{bmatrix} S(k) \begin{bmatrix} b^T \\ -I \\ A^T \end{bmatrix}^T,$$

being $\Omega(\cdot) \doteq \sum_{i=1}^{p} \mathbf{Tr} T(k-i)(X(k-i) + \hat{x}(k-i)\hat{x}(k-i)^T)$, and

$$G_2(T(k-2), \dots, T(k-p)) \doteq \begin{bmatrix} -\hat{x}(k-2)^T T(k-2) & \cdots & -\hat{x}(k-p)^T T(k-p) \\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{bmatrix},$$

$$K(S(k), T(k-2), \dots, T(k-p)) \doteq \mathbf{diag}\left(T(k-2) - \alpha_2 S(k), \dots, T(k-p) - \alpha_p S(k)\right).$$

4.4 Optimizing the size of the confidence ellipsoid

There are many ways to measure the size of an ellipsoid. We will not discuss the different possible choices: volume, diameter, etc. Here, we choose to measure the size of an ellipsoid by the sum of the squared semi-axis lengths. If the ellipsoid is described as $\mathcal{E}(X, \hat{x})$, then the size is $\mathbf{Tr}X$.

The problem of minimizing the size becomes

$$\text{minimize } \mathbf{Tr}X(k) \text{ subject to (8)}, \quad S(k) \succeq 0.$$

The above is a semidefinite program in variables $X(k), \hat{x}(k), S(k), T(k-1), \dots, T(k-p)$.

Let us discuss the complexity of our estimation method, using a *general-purpose* interior-point method for semidefinite programming, such as the one detailed in (Vandenberghe and Boyd, 1996). The LMI constraints involve two matrices, one is $S(k)$, which is of order n, the other inequality involves a matrix of order $n(p+2)+1 = O(np)$. The number of variables is $n(n+1)(p/2+1) + n = O(n^2p)$. In practice, it is observed that the number of iterations of interior-point methods for SDP is almost constant, independent of problem size (Vandenberghe and Boyd, 1996). Each iteration costs about $O(p^4n^6)$. This might seem prohibitive for very large-scale problems, but for moderate-size problems the algorithm is tractable. In addition, it is possible to exploit the structure of the problem to improve performance. Research along these lines is under way.

5. CONCLUSIONS

In this paper we introduced a new class of bounded uncertainty models. These models may be viewed as a deterministic counterpart of stochastic ARCH models, which are extensively used in financial mathematics for return and volatility modelling. The main point of the paper is that bounded uncertainty ARCH models are computationally tractable, both from the point of view of model construction (parameter fitting) and from that of set-valued forecasting. Identification of a particular model in the considered class is performed using a criterion of consistency with observed empirical data. A parallel line of research is developing the theoretical statistical foundations that lie behind the construction of this type of data-consistent models, see (Calafiore *et al.*, 2002) for further details.

REFERENCES

Avellaneda, M. and A. Parás (1996). Managing the volatility risk of portfolios of derivative securities: the Lagrangian uncertain volatility model. *Applied Mathematical Finance* **3**, 21–52.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

Boyd, S., L. El Ghaoui, E. Feron and V. Balakrishnan (1994). *Linear Matrix Inequalities in System and Control Theory*. Vol. 15 of *Studies in Applied Mathematics*. SIAM. Philadelphia, PA.

Calafiore, G., M. Campi and L. El Ghaoui (2002). Identification of reliable predictor models for unknown systems: a data-consistency approach based on learning theory. In: *Proceedings of IFAC 2002 World Congress*.

Doyle, J. (1982). Analysis of feedback systems with structured uncertainties. *IEE Proceedings* **129-D(6)**, 242–250.

El Ghaoui, L. and G. Calafiore (1999). *Worst-case simulation of uncertain systems*. pp. 134–146. Vol. 245 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag. London.

Goemans, M.X. and D.P. Williamson (1994). .878-approximation for MAX CUT and MAX 2SAT. *Proc. 26th ACM Sym. Theor. Computing* pp. 422–431.

Gouriéroux, C. (1997). *ARCH models and financial applications*. Series in Statistics. Springer. London.

Kurzhanski, A.B. and I. Valyi (1996). *Ellipsoidal Calculus for Estimation and Control*. Birkhauser. Boston.

Luenberger, David G. (1995). *Microeconomic Theory*. McGraw-Hill. New York.

Norton, J. P. and S. M. Veres (1991). Developments in parameter bounding. In: *Topics in Stochastic Systems: Modelling, Estimation and Adaptive Control* (L. Gerencser and P. E. Caines, Eds.). pp. 1137–1158. Springer-Verlag. Berlin.

Shiryaev, A.N. (1999). *Essential of Stochastic Finance*. Vol. 3 of *Advanced series on statistical science and applied probability*. World Scientific.

Vandenberghe, L. and S. Boyd (1996). Semidefinite programming. *SIAM Review* **38**(1), 49–95.

Vandenberghe, L., S. Boyd and S.-P. Wu (1998). Determinant maximization with linear matrix inequality constraints. *SIAM J. on Matrix Analysis and Applications* **19**(2), 499–533.