Mean Variance Optimization of Non–Linear Systems and Worst–case Analysis

P. Parpas[*], B. Rustem[*][†], V. Wieland [‡], S. Žaković[*]

February 8, 2006

Abstract

In this paper, we consider expected value, variance and worst–case optimization of nonlinear models. We present algorithms for computing optimal expected values, and variance, based on iterative Taylor expansions. We establish convergence and consider the relative merits of policies based on expected value optimization and worst-case robustness. The latter is a minimax strategy and ensures optimal cover in view of the worst-case scenario(s) while the former is optimal optimal expected performance in a stochastic setting. Both approaches are used with a macroeconomic policy model to illustrate relative performance, robustness and trade-offs between the strategies.

JEL Classification Codes: C61, E43

[*]Department of Computing, Imperial College, London SW7 2AZ
[†]Corresponding author. E-mail: br@doc.ic.ac.uk, tel: +44-20-7594-8345.
[‡]Goethe University Frankfurt, Frankfurt am Main, Germany

1 Introduction

Model–based policy design entails a reasonable specification of the underlying model and an appropriate characterization of the uncertainties. The latter can be an exogenous effect, parameter uncertainty, or uncertainty regarding model structure. The latter requires a setting that admits rival structures purporting to represent the same underlying system. In this paper, we consider methods that address the first two types of uncertainty.

The two approaches used are expected value optimization of nonlinear systems, and minimax, or worst–case optimization. The starting point for the former is the expected value evaluation used in [2, 14] for systems governed by parametrized feedback rules. The starting point for the the worst–case optimization approach is that of Rustem and Howe [13]. The results from both are compared in order to explore the trade–off between robustness cover and performance (as measured by the objective or cost function).

The worst–case approach to economic policy design in this paper is an application of minimax to decision making. The problem solved is the minimization of a convex (or locally convex) objective function with respect to the decision variable, and maximization of the same function with respect to the uncertainties. The uncertainties are characterized in terms of ranges in which the uncertain parameters or exogenous effects may vary.

When the cost or objective function is also convex with respect to the uncertain variables the maximum will be at the boundary of the feasible region. This may, for example, correspond to one or more vertices of the hypercube defined by the upper and lower bounds on the uncertain variables. If the objective function is concave with respect to the uncertainties, the maximum may lie anywhere within the hypercube. An advantage of the present approach is that it is straightforwardly applicable to nonlinear systems.

An alternative worst–case approach is the H^{∞} formulation (eg Basar and Bernhard [1]). The H^{∞} approach transforms the original minimax problem with box constraints, which may be convex with respect to the uncertain variables, to a concave maximization problem by an appropriate choice of a penalty parameter γ. This requires the solution of a minimax saddle point problem, convex in the minimization (i.e. policy) variables and concave in the maximization variables (i.e. uncertainties). Although the formulation is sensitive to the choice of γ, Basar and Bernhard in [1] give conditions that ensure a unique saddle point solution, thus providing a degree of robustness cover.

The distinguishing features of the mean–variance optimization in this paper are: a general approach to nonlinear dynamic systems and use of quasi–Monte Carlo (MC) simulations [16, 17] to determine the discrepancy (bias) between analytical and numerical evaluations of expectations arising from the nonlinearity of the system.

2 The Stochastic Problems

Assume that a stochastic system $f(x,v)$, is given:

$$f(x,v) = \begin{bmatrix} f_1(x,v) \\ f_2(x,v) \\ \vdots \\ f_k(x,v) \end{bmatrix}, \tag{1}$$

and let a function $F(x,v)$ be defined as follows:

$$F(x,v) = f^t(x,v)f(x,v): \quad \mathcal{R}^{n+m} \to \mathcal{R}, \tag{2}$$

$x \in \mathcal{R}^n$ and $v \in \mathcal{R}^m$. We assume that v contains noise, so $v = \overline{v} + \epsilon$, where ϵ has a normal distribution, with zero mean and Λ deviation: $\epsilon \sim \mathcal{N}(0,\Lambda)$.

The problems we consider in this paper are expected value optimization:

$$\min_x E_v(F(x,v)). \tag{3}$$

We also consider the optimization of the variance of $F(x,v)$:

$$\min_x Var_v(F(x,v)). \tag{4}$$

For non–linear models, in general, it can not be assumed that the deterministic value of the objective function is a satisfactory measure of the mean value. There are a number of studies of nonlinearity that have demonstrated the discrepancy between the two can be numerically important [4, 6, 7]. It is possible, using the Taylor series expansion, to refine the computation of $E_v(F(x,v))$ by taking into account any bias which is due to nonlinearity of the model in computing this expectation [8, 14].

Proposition 1. Let $\epsilon \in R^n, \epsilon \sim \mathcal{N}(0,\Lambda)$, and $Q \in R^{n\times n}$ a symmetric matrix. Then we have

$$E(\epsilon^t Q\epsilon) = trace(\Lambda Q).$$

Proof. Let ϵ_i denote the $i-th$ element of ϵ and Q_{ij} the $ij-th$ element of Q. Evaluating the quadratic, we obtain

$$E(\epsilon^t Q\epsilon) = E[\sum_{i,j} \epsilon_i Q_{i,j} \epsilon_j] = \sum_{i,j} Q_{i,j} E(\epsilon_i \epsilon_j) = \sum_{i,j} Q_{i,j} \Lambda_{i,j}, \tag{5}$$

from which the required result follows. ■

Proposition 2. Let $v \in R^n, v \sim \mathcal{N}(\overline{v}, \Lambda)$, and $q \in R^n$. Then

$$E_v(q^t v)^2 = (q^t \overline{v}) + q^t \Lambda q.$$

Proof. To prove the above proposition we introduce a new variable $\epsilon \sim \mathcal{N}(0, \Lambda)$, so that $v = \overline{v} + \epsilon$.

$$\begin{aligned} E_v(q^t v)^2 &= E_v(q^t v v^t q) \\ &= E_v(q^t(\overline{v} + \epsilon)(\overline{v} + \epsilon)^t q) \\ &= (q^t \overline{v})^2 + q^t \Lambda q, \end{aligned}$$

as $E_v(\epsilon) = 0$ and $E_v(\overline{v}) = \overline{v}$. ■

Proposition 3. Let $v \in R^n, v \sim \mathcal{N}(0, \Lambda)$ and Q a symmetric matrix of dimension n. Then

$$E_v[(v^t Q v)]^2 = [trace(\Lambda Q)]^2 + 2trace(\Lambda Q)^2.$$

Proof. Let the matrix $\Lambda^{1/2}$ be symmetric and $\Lambda^{1/2} \cdot \Lambda^{1/2} = \Lambda$ Furthermore, let $v = \Lambda^{1/2} b$. Thus we have:

$$E(bb^t) = I, \quad (v^t Q v) = (b^t B b),$$

where $B = \Lambda^{1/2} Q \Lambda^{1/2}$ and B is a symmetric matrix.

The components of vector b, where we denote the ith component with b_i are uncorrelated normally distributed variables and it follows from [12] that

$$E(b_i)^2 = 1, \quad E(b_i)^4 = 3, \quad \forall i.$$

Consider the transformed expression:

$$E(b^t B b)^2 = E \sum_{i,j,k,l} b_i b_j b_k b_l B_{ij} B_{kl}.$$

The only nonzero terms arise from equality of all indices or equality in pars

- $i = j, \quad k = l, \quad i \neq k;$
- $i = k, \quad j = l, \quad i \neq j;$
- $i = l, \quad j = k, \quad i = j;$
- $i = j = k = l.$

So we have

$$\begin{aligned} E(b^t B b)^2 &= \sum_{i,k,i \neq k} B_{ii} B_{kk} + \sum_{i,j,i \neq j} B_{ij}^2 \\ &+ \sum_{i,j} B_{ij} B_{ji} + 3 \sum_{i} B_{ii}^2 \\ &= \sum_{i,k} B_{ii} B_{kk} + 2 \sum_{i,j} B_{ij}^2 \qquad (6) \\ &= [trace(B)]^2 + 2 trace(B^2). \end{aligned}$$

Noting that for two square matrices D, F, $trace(DF) = trace(FD)$ we have

$$trace B = trace \Lambda^{1/2} Q \Lambda^{1/2} = trace \Lambda^{1/2} \Lambda^{1/2} Q = trace \Lambda Q,$$

$$trace B^2 = trace \Lambda^{1/2} Q \Lambda^{1/2} \Lambda^{1/2} Q \Lambda^{1/2} = trace \Lambda Q \Lambda Q = trace (\Lambda Q)^2.$$

■

2.1 Expected Value Optimization

To solve problem (3) Taylor series expansion in the neighborhood of $\overline{v}$ is used, first to compute $E_v(f_i(x,v))$, and consequently approximations for $f_i(x,v)$, and $E_v(F(x,v))$. Evaluating $f_i(x,v)$ yields:

$$\widehat{f}_i(x,v) = f_i(x,\overline{v}) + \frac{\partial f_i}{\partial v}(v-\overline{v}) + \frac{1}{2}(v-\overline{v})^t \frac{\partial^2 f_i}{\partial v^2}(v-\overline{v}) + \alpha_i(x),$$

where $\alpha_i(x)$ is evaluated to ensure that

$$E_v(f_i(x,v)) = E_v(f_i(x,\overline{v}) + \frac{\partial f_i}{\partial v}(v-\overline{v}) + \frac{1}{2}(v-\overline{v})^t \frac{\partial^2 f_i}{\partial v^2}(v-\overline{v}) + \alpha_i(x)). \quad (7)$$

The expected value on the right of (7) is evaluated as follows:

$$\begin{aligned} E_v(f_i(x,v)) &= E_v(f_i(x,\overline{v}) + \frac{\partial f_i}{\partial v}(v-\overline{v}) + \frac{1}{2}(v-\overline{v})^t \frac{\partial^2 f_i}{\partial v^2}(v-\overline{v}) + \alpha_i(x)) \\ &= f_i(x,\overline{v}) + \frac{1}{2}E_v((v-\overline{v})^t \frac{\partial^2 f_i}{\partial v^2}(v-\overline{v})) + \alpha_i(x) \quad (8) \\ &= f_i(x,\overline{v}) + \frac{1}{2} trace(\Lambda \frac{\partial^2 f_i}{\partial v^2}) + \alpha_i(x), \end{aligned}$$

which follows from Proposition 1 as $v - \overline{v} \sim \mathcal{N}(0,\Lambda)$. Thus, $E_v(f_i(x,v))$ is given by

$$f_i(x,\overline{v}) + \frac{1}{2} trace(\Lambda \frac{\partial^2 f_i}{\partial v^2}),$$

and $\alpha_i(x)$ the expected deviation of this value from $E_v(f_i(x,v))$ which, from (8), can be expressed as:

$$\alpha_i(x) = E_v(f_i(x,v)) - f_i(x,\overline{v}) - \frac{1}{2} trace(\Lambda \frac{\partial^2 f_i}{\partial v^2}). \quad (9)$$

The expectation in $\alpha_i(x)$ is estimated using MC simulation of the stochastic disturbance v.

To solve the minimization problem (3) the following expectation needs to be computed:

$$\begin{aligned} E_v(\widehat{F}(x,v)) &= E_v(\widehat{f}^t(x,v)\widehat{f}(x,v)) \\ &= E_v(\sum_{i=1}^{k}\widehat{f}_i(x,v)^2) = \sum_{i=1}^{k} E_v(\widehat{f}_i(x,v)^2). \end{aligned} \tag{10}$$

Using the Taylor series expansion (7) yields:

$$\widehat{f}_i(x,v)^2 = (f_i(x,\overline{v}) + \frac{\partial f_i}{\partial v}(v-\overline{v}) + \frac{1}{2}(v-\overline{v})^t \frac{\partial^2 f_i}{\partial v^2}(v-\overline{v}) + \alpha_i(x))^2. \tag{11}$$

Now, it is possible to compute $E_v(\widehat{f}_i(x,v)^2)$:

$$\begin{aligned} E_v(\widehat{f}_i(x,v)^2) &= \frac{\partial f_i}{\partial v}^t \Lambda \frac{\partial f_i}{\partial v} + \frac{1}{4} trace^2(\Lambda \frac{\partial^2 f_i}{\partial v^2}) \\ &+ \frac{1}{2} trace(\Lambda \frac{\partial^2 f_i}{\partial v^2})^2 + \tau_i(x)^2 \\ &+ \tau_i(x) \cdot trace(\Lambda \frac{\partial^2 f_i}{\partial v^2}), \end{aligned} \tag{12}$$

where $\tau_i(x) = f_i(x,\overline{v}) + \alpha_i(x)$.

Problem (3) can then be transformed into the following unconstrained optimization problem:

$$\min_x \sum_{i=1}^{k} E_v(\widehat{f}_i(x,v)^2). \tag{13}$$

Corollary 1. If $f_i(x,v)$ is quadratic in v, then $\alpha_i(x) = 0$, and the expected value of (2) is exactly computed by:

$$E_v(F(x,v)) = \sum_{i=1}^{k} E_v(\widehat{f}_i(x,v)^2).$$

If the problem is of higher order, then a quadratic approximation is used for minimizing expected value. An iterative approach to solving higher dimensional problems is presented below. The algorithm is based on solving the deterministic solution (for $\overline{v}$) and determining the bias $\alpha_i(x)$, the expected deviation due to the nonlinearity. It requires repeated solution of the problem as shown in Algorithm 1.

Algorithm 1: Expected Value Optimization

STEP 0: Initialization:

$l = 0$, choose x_0

STEP 1: Calculate $\alpha_i^l = \alpha_i(x_l) \forall i$, using MC simulation

STEP 2: Solve

$x_{l+1} = \arg\min_x E_v(F(x, v))$ (from (13))

STEP 3: Check for convergence:

if $\frac{\|x_{l+1} - x_l\|}{\|x_l\|} \leq \epsilon$ stop, otherwise $l = l + 1$, goto STEP 1

STEP 4: End

It is shown in Corollary 1 that $\alpha(x)$ is zero when considering model equations which are quadratic in the uncertainty v. If the uncertainty is normally distributed, the mean of the objective function can be evaluated analytically as a quartic in the uncertain variables. Algorithm 1 converges in one step in this case. However, Step 2 requires an iterative procedure for solving the optimization problem which is generally nonlinear with respect to the decision variables.

The convergence of the algorithm is tested in Step 3 to check if a fixed-point has been reached. The convergence of the algorithm is discussed below. Additionally, numerical experience has been positive. As also reported in [2, 14], even for nonlinear models, $\alpha_i(x)$ does not seem to change appreciably after the first iteration of the algorithm.

Proposition 4. Suppose that $\{x_l\}$ is a sequence generated in Step 2 of Algorithm 1, and that it remains in a compact set $\mathcal{X}$. Furthermore, suppose that either of the following two assumptions holds.

1. There exists some polynomial $p : \mathcal{R}^m \to \mathcal{R}$ such that:

$$f_i(\cdot, v) \leq p(v) \quad \forall v \in \mathcal{R}^m, i = 1, \ldots, k$$

2. Let

$$K_\infty = \{\hat{v} \in \mathcal{R}^m \mid \lim_{v \to \hat{v}} |f(\cdot, v)| = \infty\}.$$

   Suppose that $G(K_\infty) = 0$, where G denotes the Gaussian measure on $\mathcal{R}^m$.

Then any limit point of $\{x_l\}$ minimizes the expectation of $F(x, v)$ on $\mathcal{X}$.

Proof. Let x^* be any limit point of $\{x_l\}$. By the definition of $\alpha_i(x)$ in (9), it follows that the each member of the family of functions $\mathcal{F} = \{\alpha_i(x)\}$ is uniformly continuous. If either condition (1) or (2) above hold then it follows that $\alpha_i(x)$ is also G-almost surely bounded. Since $\mathcal{F}$ is a finite set it follows that the members of $\mathcal{F}$ are uniformly equi-continuous and bounded. It follows from the Arzelà-Ascoli Theorem [5] that every sequence from $\mathcal{F}$ has a subsequence which is uniformly convergent in $\mathcal{X}$. Therefore

$$\{\alpha_i(x_l)\} \rightarrow \alpha_i(x^*).$$

The result now follows from (7). ■

Remark: The conditions in Proposition 4 above are sufficiently general for most applications. If however, the function grows without bound, and this growth occurs on a set of positive measure, then one could introduce the noise in such a way so that it has its support on a compact set. In this approach we introduce the noise through an auxiliary function:

$$v = \overline{v} + g(\epsilon).$$

The auxiliary function $g : \mathcal{R}^m \rightarrow \mathcal{R}^m$ is defined as follows:

$$g(x) = \mathbf{1}_{\{x \in K\}}(x - \mu), \tag{14}$$

where:

$$\begin{aligned} \mathbf{1}_{\{x \in K\}} &= \begin{cases} 1 & \text{if } x \in K, \\ 0 & \text{otherwise.} \end{cases} \\ \mu_i &= \int_K \omega_i n(\omega) d\omega \\ n(\omega) &= \frac{\exp\left(-\frac{1}{2}\omega^T \Lambda^{-1} \omega\right)}{(2\pi)^{\frac{n}{2}} det(\Lambda)^{\frac{1}{2}}}. \end{aligned} \tag{15}$$

K is defined as the hypercube: $[-a, a]^m$, for some finite a. The derivations of this section remain largely the same so we omit the details.

2.2 Variance Optimization

When minimizing expected value performance, it is possible to consider quadratic approximations. However, when considering variance optimization, a linear approximation is the only computationally viable option. An analysis of the first order approximation is proposed in this section. The variance is given by:

$$Var_v(F(x,v)) = E_v[F(x,v) - E_v(F(x,v))]^2. \tag{16}$$

Let the model be given, as in (1):

$$f(x,v) = \begin{bmatrix} f_1(x,v) \\ f_2(x,v) \\ \vdots \\ f_k(x,v) \end{bmatrix}. \tag{17}$$

The first order Taylor series approximation of $f_i(x,v)$ in the neighborhood of $\overline{v}$ and the corresponding expectation yield:

$$\begin{aligned} \widehat{f_i}(x,v) &= f_i(x,\overline{v}) + \frac{\partial f_i}{\partial v}\epsilon + \delta_i(x) \qquad (18) \\ E(f_i(x,v)) &= f_i(x,\overline{v}) + \delta_i(x). \end{aligned}$$

As in previous section, $\delta_i(x)$ represents the expected deviation of $f_i(x,\overline{v})$ from $E_v(f_i(x,v))$ which, from (18), can be expressed as:

$$\delta_i(x) = E_v(f_i(x,v)) - f_i(x,\overline{v}), \tag{19}$$

and the expectation in $\delta_i(x)$ is estimated using a quasi–Monte Carlo simulation of the stochastic disturbance v.

Expected value of the quadratic objective function $F(x,v)$ is evaluated as follows:

$$E(\widehat{F}(x,v)) = E\sum_{i=1}^{k}\widehat{f}_i(x,v)^2 = \sum_{i=1}^{k}E(\widehat{f}_i(x,v)^2), \tag{20}$$

where $E(f_i(x,v)^2)$ is computed using the expansion in (18):

$$\begin{aligned} E(\widehat{f}_i(x,v)^2) &= E(f_i(x,\overline{v}) + \frac{\partial f_i}{\partial v}\epsilon + \delta_i(x))^2 \\ &= (f_i(x,\overline{v}) + \delta_i(x))^2 + \frac{\partial f_i}{\partial v}^t \Lambda \frac{\partial f_i}{\partial v}. \end{aligned} \tag{21}$$

The problem of minimizing the variance is formulated as:

$$\min_x Var_v(\widehat{F}(x,v)) = \min_x \sum_{i=1}^{k} Var_v(\widehat{f}_i(x,v)^2), \tag{22}$$

where

$$\begin{aligned} Var_v(\widehat{f}_i(x,v)^2) &= E\Big(\frac{\partial f_i}{\partial v}\epsilon\Big)^4 + 4\varphi(x)^2\frac{\partial f_i}{\partial v}^t\Lambda\frac{\partial f_i}{\partial v} - \Big(\frac{\partial f_i}{\partial v}^t\Lambda\frac{\partial f_i}{\partial v}\Big)^2 \\ &= trace(\Lambda Df_i(x))^2 + 2trace^2(\Lambda Df_i(x)) \\ &+ \ 4\varphi(x)^2\frac{\partial f_i}{\partial v}^t\Lambda\frac{\partial f_i}{\partial v} - \Big(\frac{\partial f_i}{\partial v}^t\Lambda\frac{\partial f_i}{\partial v}\Big)^2, \end{aligned} \tag{23}$$

$\varphi(x) = \delta_i(x) + f_i(x,\overline{v})$, and $Df_i(x) = \frac{\partial f_i}{\partial v}\frac{\partial f_i}{\partial v}^t$.

Corollary 2. If $f_i(x,v)$ is linear in v then $\delta_i(x) = 0$, and $Var_v(F(x,v))$ is exactly computed by:

$$Var_v(F(x,v)) = \sum_{i=1}^{k} Var_v(\widehat{f}_i(x,v)^2).$$

An iterative approach to solving higher dimensional problems is presented in Appendix A. If the problem is of higher order, then a linear approximation is used for minimizing variance.

3 The Worst–case Approaches

In the previous two sections we were interested in the optimization of the expected value or variance of the objective function. We now turn our attention to a different approach: worst–case analysis. The latter type of analysis has a game–theoretic interpretation. The first player is the decision–maker, choosing the decision vector x. The second player is nature, and is assumed to be antagonistic to the decision maker, nature selects the realizations of the random variables. Therefore, the aim of worst–case analysis is to minimize the objective function with respect to the worst possible outcome of the uncertain variables v. In this section, two worst–case approaches are considered, namely the minimax and H^{∞} approach.

3.1 The Minimax Approach

According to the framework described above, the optimization problem we consider in this section is given by:

$$\begin{aligned} \min_{x} \quad \max_{v} \quad & F(x, v), \\ s.t. \quad & \overline{v} - \Delta \leq v \leq \overline{v} + \Delta, \quad \Delta > 0. \end{aligned} \tag{24}$$

Due to the hypercube constraining (24), the problem above is referred to as box-constrained.

Let

$$\Phi(x) = \max_{\overline{v} - \Delta \leq v \leq \overline{v} + \Delta} F(x, v), \tag{25}$$

for all x. We call $\Phi(x)$ the max–function. Therefore, (24) can be written as

$$\min_x \Phi(x). \tag{26}$$

To solve (26) a quasi–Newton algorithm is used. The algorithm generates a descent direction based on a subgradient of $F(x, .)$ and uses an approximate Hessian (H_k) in the presence of possibly multiple maximizers of (25) as well as a step size strategy that ensures sufficient decrease in $\Phi(x)$ at each iteration (Rustem and Zakovic [15]).

Problem (26) poses several difficulties:

- $\Phi(x)$ is in general continuous but may have kinks, so it might not be differentiable. At a kink the maximizer is not unique and the choice of subgradient to generate a search direction is not simple;
- $\Phi(x)$ may not be computed accurately as it would require infinitely many iterations of an algorithm to maximize $f(x, y)$;
- In (26) a global maximum is required in view of possible multiple solutions. The use of a local maximum cannot guarantee a monotonic decrease in $\Phi(x)$.

Full minimax algorithms and applications to a number of problems in engineering, finance and macroeconomics are presented in [13, 15, 19, 20]. The issue of global maxima is further considered in Section 4.2.

3.2 The H^∞ Approach

Another approach to robust design, with a minimax origin, is the H^∞ framework (Basar and Bernhard [1]). In the minimax approach the uncertainty v

is allowed to take arbitrary values from the feasible region, regardless of how low probability of those values occurring is. In other words, nature's strategy is to place all its mass on the worst–case scenario. The H^∞ approach, and its precursors (Bernhard and Bellec [3], Jacobson[10]) take this into account by including a penalty term within a linear–quadratic control framework. The two papers show that the importance and robustness of minimax and worst–case design has been recognized for more than 30 years.

Consider the following problem

$$\begin{aligned} \min_x \max_v \quad & F(x, v) \\ s.t. \quad & \|v - \overline{v}\|_2^2 \leq C. \end{aligned} \tag{27}$$

If F is convex in x and convex in v, the solution of this problem lies on the boundary of the constraint $\|v - \overline{v}\|_2^2 \leq C$. Furthermore, generally, there are multiple maximizers. The constraint is implicitly imposed using a penalty formulation that discourages its transgression. Hence, the H^∞ formulation is given by

$$\min_x \max_v F(x, v) - \gamma^2 \|v - \overline{v}\|_2^2, \tag{28}$$

where $\gamma > 0$. For small values of γ, (28) is convex in x and convex in v. Hence, it can have multiple maxima. As γ increases, the augmented objective function (28) becomes concave in v at some value of γ. It is this value of γ that H^∞ seeks. The approach is sensitive to the choice of γ but yields a robust solution that is also a saddle point as the transformed objective is convex in x and concave in v for larger γ. The choice of γ is further discussed in Section 4.2.

3.3 Robustness and Optimality of Minimax

Robustness and the price paid for this desirable property, has been the topic of interest for a number of years [10]. For both, the minimax and H^∞ formulation, robustness is ensured by an optimality condition. Let x^*, v^* solve (24). Then we have

$$F(x^*, v^*) \geq F(x^*, v), \text{ for all feasible } v,$$

and let x^{**}, v^{**} solve (28). Then

$$F(x^{**}, v^{**}) - \gamma^2 \|v^{**} - \overline{v}\|_2^2 \geq F(x^{**}, v) - \gamma^2 \|v - \overline{v}\|_2^2, \text{ for all feasible } v.$$

The above inequalities simply state the optimality of v^*, v^{**} for the corresponding problem. However, they encompass the robustness of minimax in that performance is assured to improve if the worst–case v^*, or v^{**} does not happen.

Similarly, under the same assumptions

$$F(x^*, v^*) \leq F(x, v^*), \text{ for all feasible } x,$$

and

$$F(x^{**}, v^{**}) - \gamma^2 \|v^{**} - \overline{v}\|_2^2 \leq F(x, v^{**}) - \gamma^2 \|v^{**} - \overline{v}\|_2^2, \text{ for all feasible } x.$$

This is illustrated in Section 4.3 where we compute minimax, x^*, v^*, and expected value optima, x_e, and confirm the inequality

$$F(x^*, v^*) \leq F(x_e, v^*).$$

4 Numerical Results

One can present arguments for and against expected value optimization, and similarly for worst–case analysis. Using the methods to solve real world problems is bound to give more insight into the usefulness and properties of the two frameworks adumbrated in previous sections. In this section we will present and compare results obtained with the two different approaches:

- Worst–case analysis using the minimax formulation:

$$\begin{aligned} \min_x \quad \max_v \quad & F(x, v), \\ s.t. \quad & \overline{v} - \sigma_v \leq v \leq \overline{v} + \sigma_v. \end{aligned}$$

- Minimization of expected value performance:

$$\begin{aligned} \min_x \quad & E_v(F(x, v)) \\ s.t. \quad & v \sim \mathcal{N}(\overline{v}, \Lambda). \end{aligned}$$

4.1 A Model of the Economy

In a recent paper, Orphanides and Wieland [11] use a simple macroeconomic model of inflation, output and interest rates to investigate different motives for inflation point versus inflation zone targeting. In the first case, the policy-maker varies short-term nominal interest rates in order to stabilize inflation around a target point. In the second case, the emphasis is on containing inflation within a target range. Inflation point targeting arises naturally in linear models of the economy with a quadratic loss function for the policymaker (the L–Q model in [11]). Orphanides and Wieland show that inflation zone targeting may be motivated by a non-linear, or more precisely, zone-linear Phillips curve relationship between the change in inflation and the output gap (the ZL–Q model in [11]).

In the minimalist macro model of [11], the two key variables for the policy decision process are inflation and output. The policy instrument is the short term nominal interest rate. The dynamic structure of the model is represented by a single lag of inflation in the Phillips curve, and a single lag of the output gap in the aggregate demand equation. It is appropriate, therefore, to interpret the length of a period to be rather long, say half a year to a year.

In every period, the policymaker sets the nominal interest rate, R, with the objective to maintain inflation π, close to a desired target and output close to the economy's natural level. To describe the policymaker's welfare loss during a period t, a per–period loss function is specified:

$$l_t = l(\pi_t, y_t).$$

Assuming that the policymaker discounts the future with a fixed factor β, we can view the objective in period t as to minimize the expected discounted sum of future per–period losses from $t + 1$ onwards:

$$\min_r E\{\sum_{t=1}^{\infty} \beta^{t-1} l_t\}. \tag{29}$$

The per–period loss facing the policymaker in period $t + 1$, l_{t+1} can be expressed as a weighted average of the deviation of inflation π from its desired target π^* and the output deviation from the economy's natural level y.

$$l_{t+1} = \omega(\pi_{t+1} - \pi^*)^2 + (1-\omega)y_{t+1}^2, \quad \omega \in (0,1). \tag{30}$$

The following two equations describe the evolution of the economy:

$$\begin{aligned} \begin{bmatrix} \pi_t \\ y_t \end{bmatrix} &= \begin{bmatrix} 1 & \alpha\rho \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} -\alpha\xi \\ -\xi \end{bmatrix} r_t \\ &+ \begin{bmatrix} \alpha\delta + \alpha u_t + e_t \\ \delta + u_t \end{bmatrix}, \end{aligned} \tag{31}$$

where e_t and u_t are normally distributed, zero–mean shocks:

$$u_t, e_t \sim \mathcal{N}(0, \Lambda), \quad \forall t. \tag{32}$$

The objective function is defined in terms of a sum of per–period losses l_t:

$$F(r,v) = \sum_{t=1}^{\infty} \beta^{t-1} l_t. \tag{33}$$

An alternative approach, which could be used in this framework, is that of Tetlow and von zur Muehlen [18]. In their approach (also Hansen and Sargent in [9]) the policymaker chooses the parameters x_1 and x_2 of the feedback law:

$$r_t = x_1\pi_{t-1} + x_2 y_{t-1}, \tag{34}$$

to minimize welfare losses that are maximized over w_t. This rule is referred to as a feedback rule.

The problem can be formulated as:

$$\min_{x_1, x_2} E_v(F(x,v)), \tag{35}$$

where E_v denotes the expectation computed for uncertain variables v, the objective function F is given by (33), the constraints on the systems are given by the model (31) and the feedback law given by (34).

Let

$$f_1(x,v) = \begin{bmatrix} \beta^{\frac{1}{2}}(\pi_1 - \pi^*) \\ \beta^{\frac{2}{2}}(\pi_2 - \pi^*) \\ \vdots \\ \beta^{\frac{T}{2}}(\pi_T - \pi^*) \end{bmatrix}, \quad f_2(x,v) = \begin{bmatrix} \beta^{\frac{1}{2}} y_1 \\ \beta^{\frac{2}{2}} y_2 \\ \vdots \\ \beta^{\frac{T}{2}} y_T \end{bmatrix}, \tag{36}$$

then, the objective function can be formulated as:

$$F(x,v) = w f_1^t f_1 + (1-w) f_2^t f_2, \tag{37}$$

so the problem becomes:

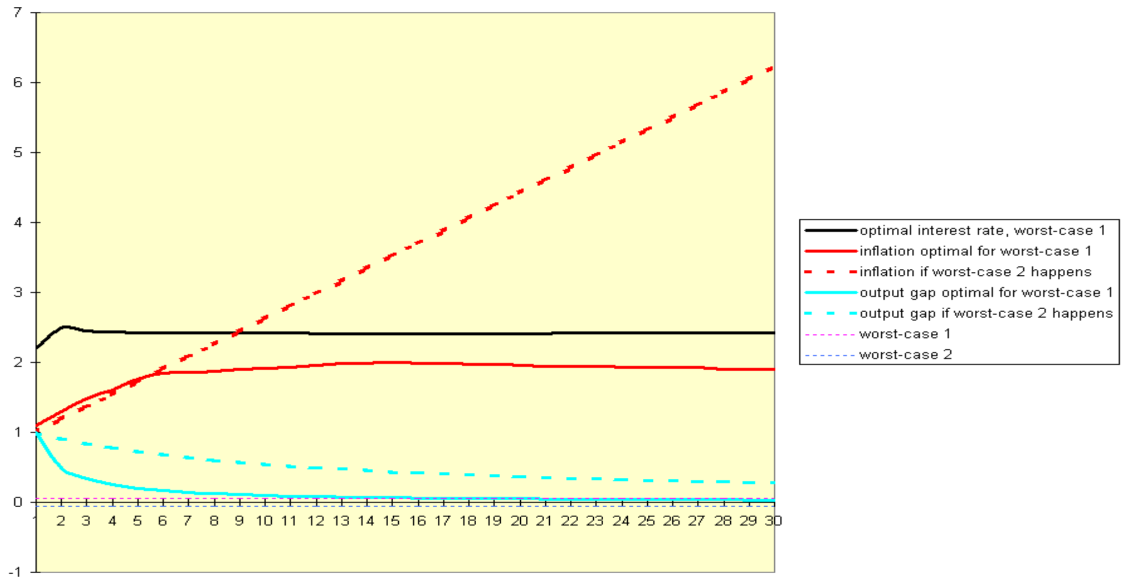
$$\min_x \{ w E_v(f_1^t(x,v) f(x,v)) + E_v(f_2^t(x,v) f_2(x,v)) \}. \tag{38}$$

Therefore, the expectation $E_v F(x,v)$ can be calculated as the sum of expectations of quadratic functions $(\pi_{t+1} - \pi^*)^2$ and y_{t+1}^2. ¿From (12) the expectations for each time period t that appear in the sum can be calculated, allowing for the bias to be evaluated with increased accuracy.

4.2 Observations on Minimax

The importance of identifying all global maxima is illustrated in Figure 1, which is based on the economic model introduced in Section 4.1. As mentioned before, the model consists of three variables - interest rate (the decision variable), output gap, and inflation (uncertain variables – contain random shocks). Figure 1 shows the behavior when the decision (interest rate) is

Figure 1: Optimizing in view of worst–case 1 only and cross evaluation of performance if worst–case 2 is realized.



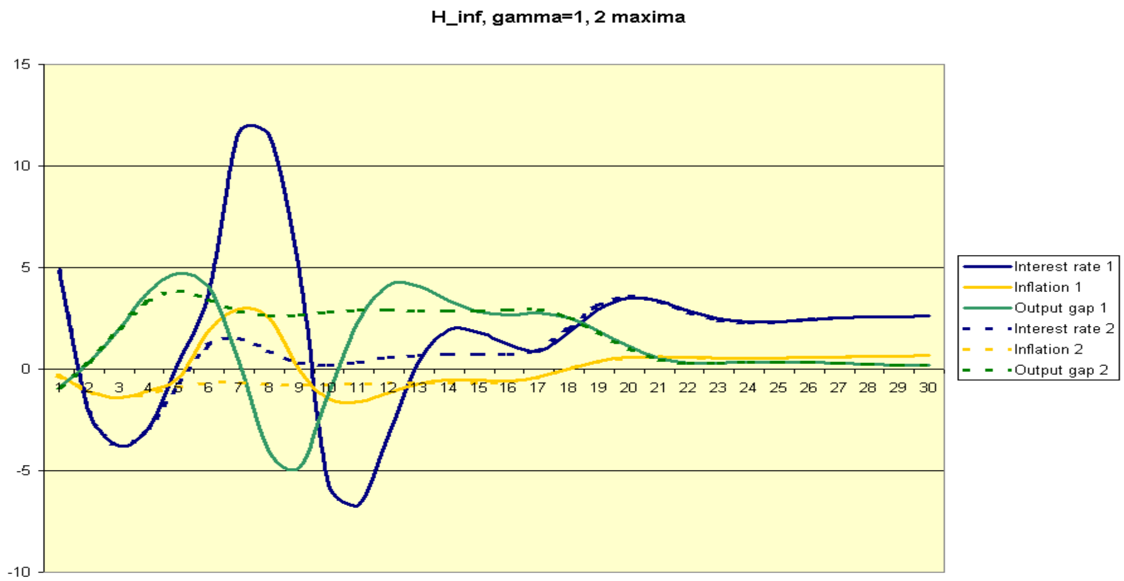
optimized with respect to one maximizer (worst–case realization of shocks) only.

If worst–case 1 is realized then inflation stays close to the given target of 2%. However if worst–case 2 is realized then inflation rises to more than 6 %, three times greater than the desired target. A similar result applies for output gap.

In the H^∞ formulation in Section 3.2, we note the importance of the choice of parameter γ. The results for various choices of γ are presented in Figures 2, 3 and 4. In the first, $\gamma = 1$, two distinctive maxima are observed (i.e. two worst–case scenarios) of interest rate, inflation, and output gap.

In Figure 3, the penalty parameter is increased to $\gamma = 2$. There are

Figure 2: Inflation patterns for $\gamma = 1$. Two maxima encountered.



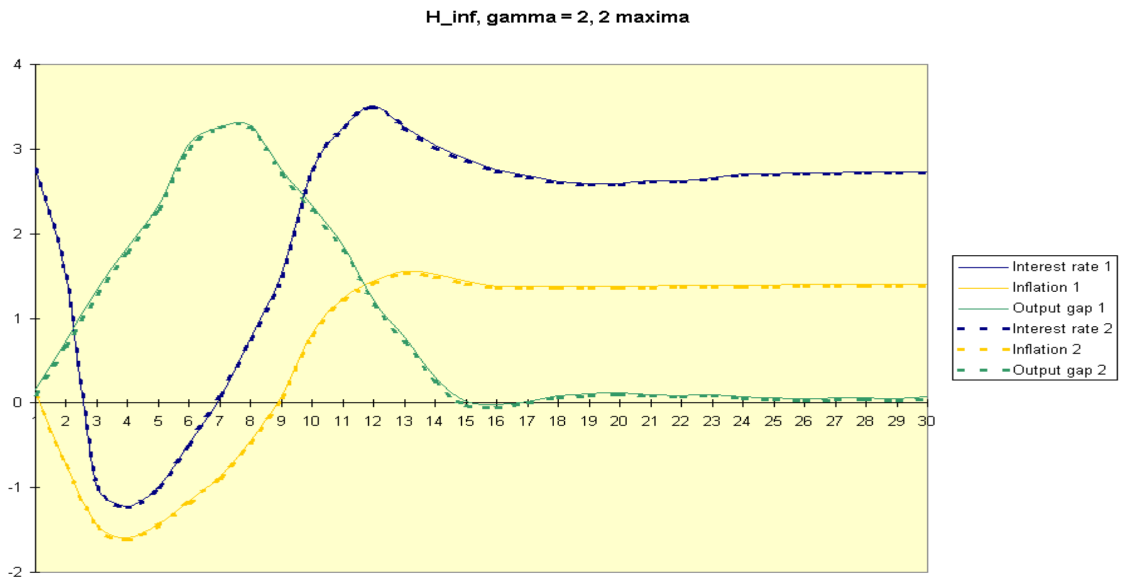
still two maxima, but it can be observed that they are much closer, almost identical.

In Figure 4, with $\gamma = 3$, the uncertainties are forced to zero and we have a saddle point. One maximum is observed for interest rate, inflation and output gap.

The advantage of the minimax approach is that it provides a worst–case value of v and enables us to evaluate the worst–case scenario. On the other hand, H^{∞} ensures robustness of the strategy from deviation of v away from its nominal value of $\overline{v}$ and as a saddle point problem it is simpler to compute (see [19]).

The robustness of minimax is illustrated in Figure 5, where there are two

Figure 3: Inflation patterns for $\gamma = 2$. Two maxima encountered.



possible worst–case scenarios (v_1^* and v_2^*), represented with two paths with the highest function values ($F(x^*, v_1^*)$ and $F(x^*, v_2^*)$). All the other paths on the graph represent different, randomly chosen scenarios and it can be seen that function values in all the other cases are significantly smaller.

4.3 Computational Experiments

The results that follow are obtained for $\overline{\pi} = 0, t = 20$, $\beta = 0.9$ and the same weight $w = \frac{1}{2}$ for both inflation and output gap. For the model parameters the estimates obtained by Orphanides and Wieland [11] are used. These estimates are summarized in Table 1. Only the estimates from the first column (Euro Area 1976–1998) are used for the numerical solutions.

There are three sets of results in Table 2, corresponding to different

Figure 4: Inflation patterns for $\gamma = 3$. One maximum encountered.

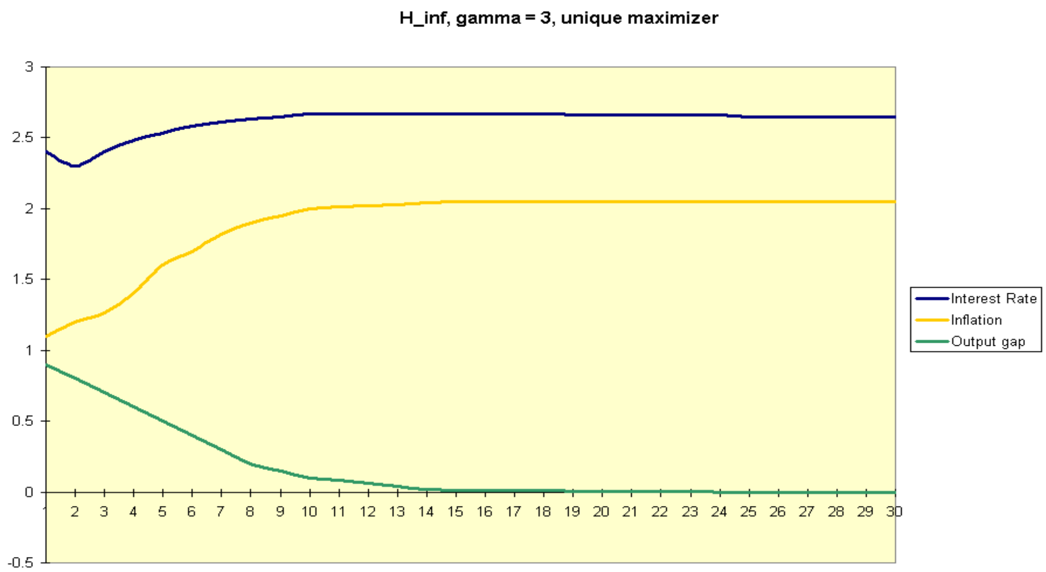
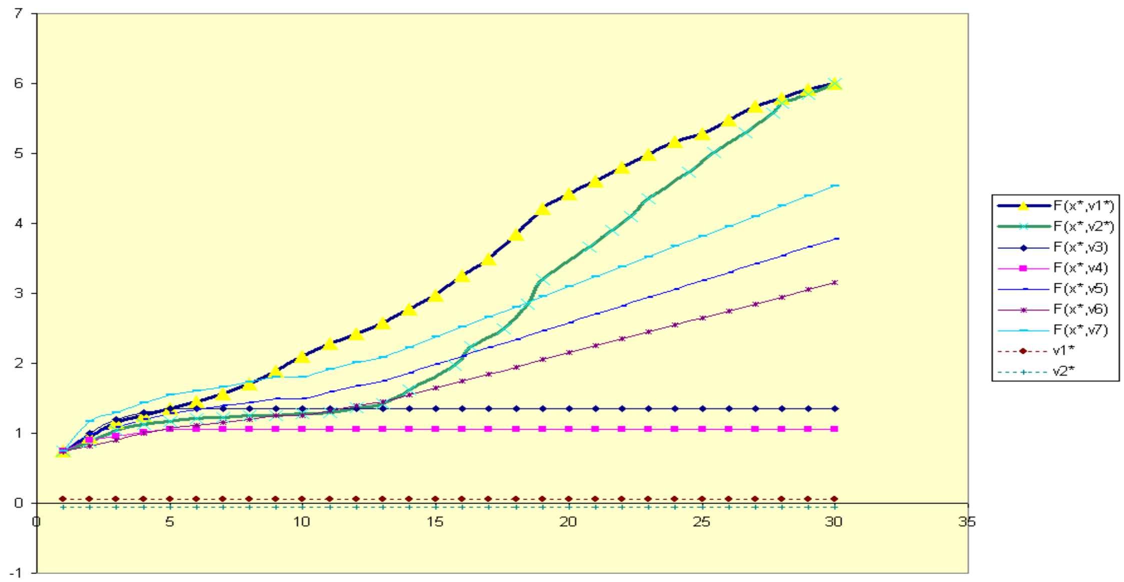


Table 1: State Equation Parameters

	Euro Area	United States		
	(OECD) 1976-1998	(OECD) 1976-1998	(CBO) 1976-1998	(CBO) 1960-1998
δ	1.07	1.03	0.54	0.64
ρ	0.77	0.47	0.64	0.63
ξ	0.40	0.32	0.23	0.23
σ_u	0.84	1.51	1.62	1.80
α	0.34	0.39	0.31	0.31
σ_e	0.96	0.85	0.89	1.06

bounds on the uncertainties. As the shocks (uncertainties u, e) are additive in the model, the feedback rules are the same for all three cases. What changes is the function value, which increases with the increment of the boundaries on uncertainties. We also report the number of worst–cases observed at each

Figure 5: Noninferiority of minimax: performance improves if the two worst cases $v_1^*$and v_2^* do not materialize. Hence, $F(x^*, v_1^*) = F(x^*, v_2^*) \geq F(x^*, v_i)$, for randomly generated $v_i, i = 3, ..., 7$.



computation.

Table 2: Linear Model - Minimax

$\sigma_u = 0.84$, $\sigma_e = 0.96$					
bounds		x_1	x_2	$F(x, v)$	maxima
$\frac{1}{2}\sigma_u$	$\frac{1}{2}\sigma_e$	5.217	1.873	10.943	2
σ_u	σ_e	5.217	1.873	43.772	2
$\frac{3}{2}\sigma_u$	$\frac{3}{2}\sigma_e$	5.217	1.873	101.252	2

In Table 3 results of minimizing the expected value are presented. A similar conclusion can be drawn: the optimal decision is the same due to the shocks appearing linearly in the model, and also the expected loss increases as the uncertainty increases.

Table 3: Linear Model - Expected values

distribution		x_1	x_2	$E(F(x,v))$
$\mathcal{N}(0,\frac{\sigma_u}{4}^2)$	$\mathcal{N}(0,\frac{\sigma_e}{4}^2)$	1.857	1.930	4.013
$\mathcal{N}(0,\sigma_u^2)$	$\mathcal{N}(0,\sigma_e^2)$	1.857	1.930	16.053
$\mathcal{N}(0,(\frac{3}{2}\sigma_u)^2)$	$\mathcal{N}(0,(\frac{3}{2}\sigma_e)^2)$	1.857	1.930	36.119

It can be observed from the results that expectation of the loss is always lower than the worst–case. Results of cross evaluation are presented in Table 4. We evaluate the consequence of applying the expected value optima (corresponding to different levels of uncertainty) when the worst–case scenario is realized. Also, the consequence of adopting the worst–case optima (corresponding to different bounds) in view of stochastic uncertainty is evaluated.

Table 4: Cross Evaluation

Minimax Optimum				
bounds	x_1	x_2	worst–case optimum	exp. val. of mmx
$\frac{1}{2}\sigma_u, \frac{1}{2}\sigma_e$	5.217	1.873	10.943	6.583
σ_u, σ_e	5.217	1.873	43.772	26.334
$\frac{3}{2}\sigma_u, \frac{3}{2}\sigma_e,$	5.217	1.873	101.252	72.638
Minimized Expectation				
distribution	x_1	x_2	exp. val. optimum	worst–case of exp. val. opt.
$\mathcal{N}(0,\frac{\sigma_u^2}{4}),\mathcal{N}(0,\frac{\sigma_e}{4}^2)$	1.857	1.930	4.013	12.564
$\mathcal{N}(0,\sigma_u^2),\mathcal{N}(0,\sigma_e^2)$	1.857	1.930	16.053	66.543
$\mathcal{N}(0,(\frac{3\sigma_u}{2})^2),\mathcal{N}(0,(\frac{3\sigma_e}{2})^2)$	1.857	1.930	36.119	129.733

We compare adopting the worst–case feedback rule in the stochastic framework and the expected value optimization feedback rule when the worst–case is realized. The expected performance of the former (completed using

MC simulation) is observed to be much better than the performance of the latter (for example 6.583 is the expectation, while the worst–case value is 10.943).

The situation rapidly changes when the feedback rules obtained from minimizing expectation are used. In case when such rules are used and the worst–case scenario happens, the loss could increase up to 60% (from 43.772 to 66.543). Therefore, this brings us to the main conclusion that, although the expected value optimization performs better on average, minimax optimization guards against the worst possible scenarios and provides the upper bound for (in this case) loss function. Performance is guaranteed for the worst–case and will improve if any scenario, other than the worst–case, is realized.

5 Conclusions

Methods for mean variance and worst–case optimization of nonlinear models have been presented. Algorithms for computing optimal expected value and variance based on iterative Taylor expansions have been developed and compared with a minimax algorithm for computing robust policies.

To compare results a simple macroeconomic model of inflation, output and interest rates due to Orphanides and Wieland [11] was used. The results presented in Section 4.3 showed that, although the expected value optimization performed better on average, the worst–case optimal strategy provided robust solutions, that performed much better under the worst–case scenarios (two have been found for all of the three experiments) and optimal cross evaluation of worst–case scenarios for expected value strategy indicates that performance deterioration for the latter could be a serious issue. The importance of finding all worst–cases, together with robustness issues of the minimax strategy has been emphasized in Section 4.2.

The H^∞ method as an alternative minimax approach has also been considered. This method is very sensitive to the choice of the penalty parameter γ, which we illustrated in Figures 2, 3 and 4.

Acknowledgments

This research was supported by two EPSRC grants – GR/T02560/01 and EP/C513584/1. The authors are grateful to an anonymous referee for helpful comments and criticisms.

References

[1] Basar, T. and Bernhard, P., H^∞–Optimal Control and Related Minimax Design Problems, Birkhauser, Boston, 1991.

[2] Becker, R. G., Hall, S., and Rustem, B., "Robust optimal decisions with stochastic nonlinear economic systems", Journal of Economic Dynamics & Control, vol. 18, pp 125–148, 2000.

[3] Bernhard, P., and Bellec, G., "On the evaluation of worst case design with an application to the quadratic synthesis technique", in Guardabassi, G., Locatelli, A., and Rinaldi, S., " Sensitivity, adaptivity and optimality : proceedings of the third IFAC symposium", June 18-23, 1973, Ischia, Italy.

[4] Chow, G. and Corsi, P. (eds.), Evaluating the Reliability of Macroeconomic Models, Wiley, New York, 1982.

[5] Dudley, R.M., "Real analysis and probability", Cambridge University Press, 2002.

[6] Fair, R. C., Specification Estimation and Analysis of Macroeconometric Models, Harvard University Press, Cambridge, MA, 1984.

[7] Hall, S. G. and Henry, S. G. B., Macroeconomic Modelling, North–Holland, Amsterdam, 1988.

[8] Hall, S. G. and Stephenson, S. J.,"Optimal control of stochastic non–linear models", in N. Christodoulakis (ed.) – Dynamic modelling & control of national economies, Pergamon Press, Oxford, 1990.

[9] Hansen, L. and Sargent, T. J., and Tallarini, T. D.,"Robust permanent income and pricing", Review of Economic Studies 66, pp. 873–907, 1999.

[10] Jacobson, D. H., "New interpretations and justifications for worst case min–max design of linear control systems", in Guardabassi, G., Locatelli, A., and Rinaldi, S., " Sensitivity, adaptivity and optimality : proceedings of the third IFAC symposium", June 18-23, 1973, Ischia, Italy.

[11] Orphanides, A. and Wieland, V., "Inflation Zone Targeting", European Economic Review, pp. 1351–1387, vol. 44, 2000.

[12] Rustem, B., "Algorithms for nonlinear programming and multiple–objective decisions", Chichester, Wiley; 1998.

[13] Rustem, B., and Howe, M.A., "Algorithms for Worst-case Design with Applications to Risk Management', Princeton University Press, 2002.

[14] Rustem, B., "Stochastic and robust control of nonlinear economic systems", European Journal of Operational Research, vol. 73, pp. 304–318, 1994.

[15] Rustem, B., and Zakovic, S., "An interior point algorithm for continuous minimax", submitted to Jota, 2005.

[16] Sobol I. M., "On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals", Comput. Math. Phys., vol. 7, pp. 86–112, 1967.

[17] Sobol, I. M., "On Quasi-Monte Carlo Integrations", Mathematics and Computers in Simulation, vol. 47, pp 103–112, 1998.

[18] Tetlow, R. J. and von zur Muehlen, P., "Robust Monetary Policy with Misspecified Models: Does Model Uncertainty Always Call For Attenuated Policy?", Journal of Economic Dynamics & Control, vol. 25, pp. 911–949, 2001.

[19] Zakovic, S., Pantelides C. C., and Rustem, B., "An Interior Point Algorithm for Computing Saddle Points of Constrained Continuous Minimax", Annals of Operations Research 99, pp 59–77, 2000.

[20] S. Zakovic, B. Rustem and V. Wieland, "A Continuous Minimax Problem and its Application to Inflation Targeting", in G. Zaccour(Ed) - Decision and Control in Management Science, Kluwer Academic Publishers, 2002.

A The Algorithm for Variance Optimization

As in case of expected value optimization, the algorithm is based on solving the deterministic solution (for $\overline{v}$) and determining the bias $\delta_i(x)$ (the expected deviation due to the nonlinearity). It requires repeated solution of the problem as shown in Algorithm 2.

Algorithm 2: Variance Optimization

STEP 0: Initialization:

$l = 0$, choose x_0

STEP 1: Calculate $\delta_i^l = \delta_i(x_l) \forall i$, using MC simulation

STEP 2: Solve

$x_{l+1} = \arg\min_x Var_v(F(x, v))$ (from (22))

STEP 3: Check for convergence:

if $\frac{\|x_{l+1} - x_l\|}{\|x_l\|} \leq \epsilon$ stop, otherwise $l = l + 1$, goto STEP 1

STEP 4: End

This algorithm converges in one step for a model linear in uncertainties and a quadratic objective. Corollary 2 ensures that $\delta_i(x) = 0$ in this case and the variance can be evaluated analytically as a function of the decision variables.